

QY 241 RTLNSTSPFBEKLEAETKTLHSLGTLAOLFSMSMSMDROEVLINVNSSSTOI 300
 Db 241 RTLNSTSPFBEKLEAETKTLHSLGTLAOLFSMSMSMDROEVLINVNSSSTOI 300
 QY 301 YQASRIYCGHPEGGLKIKSLNWEEDNNYKALFGNGTEDEAETFDNSTPYCNDLAK 360
 Db 301 YQASRIYCGHPEGGLKIKSLNWEEDNNYKALFGNGTEDEAETFDNSTPYCNDLAK 360
 QY 361 NLESPLRIIMKALPLLVKILYTPDTPATROVMAEVNKTFOELAVFHDLSGMEELS 420
 Db 361 NLESPLRIIMKALPLLVKILYTPDTPATROVMAEVNKTFOELAVFHDLSGMEELS 420
 QY 421 PKITPMESQMDLVRLNLDSDRDHFWEOQLDMDTADIDYAFIAPLAKHEDVQSSNGS 480
 Db 421 PKITPMESQMDLVRLNLDSDRDHFWEOQLDMDTADIDYAFIAPLAKHEDVQSSNGS 480
 QY 481 VYTWREAFNETNOARTISRMECVNLKLEPATEVWLINXSMELDERKFMAGIYFTG 540
 Db 481 VYTWREAFNETNOARTISRMECVNLKLEPATEVWLINXSMELDERKFMAGIYFTG 540
 QY 541 ITPGSIELPHYKKTIRADIDNVERTNKIKOGYMDGPRADEEMRYVWGCFALDQV 600
 Db 541 ITPGSIELPHYKKTIRADIDNVERTNKIKOGYMDGPRADEEMRYVWGCFALDQV 600
 QY 601 EOAIIIRVLNTEKKTGVYMOOMKPCYVDIDILRVMSRSMPLFMTLAMIYSAVIKIGIV 660
 Db 601 EOAIIIRVLNTEKKTGVYMOOMKPCYVDIDILRVMSRSMPLFMTLAMIYSAVIKIGIV 660
 QY 661 YEKEARLEKTRIMGLDINSILFESLIPPLVSAGLVILKGNLLEPSPDSVAVF 720
 Db 661 YEKEARLEKTRIMGLDINSILFESLIPPLVSAGLVILKGNLLEPSPDSVAVF 720
 QY 721 FLVSFAVVTILOCFLISTESRANLAAACGIIYFTLYLPYLCVAMODYVGTIKIEAS 780
 Db 721 FLVSFAVVTILOCFLISTESRANLAAACGIIYFTLYLPYLCVAMODYVGTIKIEAS 780
 QY 781 LLSVAFGCEYFALFEBEGIGYQMDNLEESVEDEGFLUTSISMLFDFFLYGVMT 840
 Db 781 LLSVAFGCEYFALFEBEGIGYQMDNLEESVEDEGFLUTSISMLFDFFLYGVMT 840
 QY 841 YIEAFPGOYCIIPRWYFPCRTKYWFEESDEKSHPSNOKRMEIICOMESEPTHLKGV 900
 Db 841 YIEAFPGOYCIIPRWYFPCRTKYWFEESDEKSHPSNOKRMEIICOMESEPTHLKGV 900
 QY 901 IONLVKVRDGMKAVDGLALNIEGQITTSFLGNAGKTTINSILTLGPPPSGAYIL 960
 Db 901 IONLVKVRDGMKAVDGLALNIEGQITTSFLGNAGKTTINSILTLGPPPSGAYIL 960
 QY 961 GMDISEKSTIRONLGVCPONHVLFDMLTVEHIMFYARLKGLEKRVKAKEMOMALDVG 1020
 Db 961 GMDISEKSTIRONLGVCPONHVLFDMLTVEHIMFYARLKGLEKRVKAKEMOMALDVG 1020
 QY 1021 LPSKSLKSTQSLSGMOKRISVALAFYVGSKVYILDEPTAGVPSYRSGIWEILLKXRO 1080
 Db 1021 LPSKSLKSTQSLSGMOKRISVALAFYVGSKVYILDEPTAGVPSYRSGIWEILLKXRO 1080
 QY 1081 GRITIIISTHMBADVLGRIRAIISHGKLCYVGSLEFLKNOLGTYITLVKKNVSSLS 1140
 Db 1081 GRITIIISTHMBADVLGRIRAIISHGKLCYVGSLEFLKNOLGTYITLVKKNVSSLS 1140
 QY 1141 SCRNSSSTVYLKEDSVQSSSDAGLSGDSHSDTLTIDVSAISNLIRKHVSEALVBDI 1200
 Db 1141 SCRNSSSTVYLKEDSVQSSSDAGLSGDSHSDTLTIDVSAISNLIRKHVSEALVBDI 1200
 QY 1201 GHELTYVLPYEAKEGAFVELHEIDDRISDLSISYIGSETTLEPIKVAEBEGVDAE 1260
 Db 1201 GHELTYVLPYEAKEGAFVELHEIDDRISDLSISYIGSETTLEPIKVAEBEGVDAE 1260
 QY 1261 TSOGTLPARNRRAFGQSCLEPTEDDAADPNDSIDPESHETDLSGMDKGSYOVK 1320
 Db 1261 TSOGTLPARNRRAFGQSCLEPTEDDAADPNDSIDPESHETDLSGMDKGSYOVK 1320
 QY 1321 GWKLTQOOFVALLMKRLLIARRSKGFFAOIYLPANFVCIALVPSLIVPFGKYSLELO 1380

Db 1381 GWKLTQOOFVALLMKRLLIARRSKGFFAOIYLPANFVCIALVPSLIVPFGKYSLELO 1380
 QY 1381 GWKLTQOOFVALLMKRLLIARRSKGFFAOIYLPANFVCIALVPSLIVPFGKYSLELO 1380
 Db 1381 GWKLTQOOFVALLMKRLLIARRSKGFFAOIYLPANFVCIALVPSLIVPFGKYSLELO 1380
 QY 1441 VPOTMDLPONGNMTQONPSPACQSSDKIKKMLPVCPPGAGGLPPQOKQNTADILDL 1500
 Db 1441 VPOTMDLPONGNMTQONPSPACQSSDKIKKMLPVCPPGAGGLPPQOKQNTADILDL 1500
 QY 1501 TGRNLSDTLACTYVOIILAKSLKKNLWNEFRYGGFSLGVSNTQALPPSGEVDATOKMK 1560
 Db 1501 TGRNLSDTLACTYVOIILAKSLKKNLWNEFRYGGFSLGVSNTQALPPSGEVDATOKMK 1560
 QY 1561 HLKAKSSADRFILSGRPMTGDLTRNNKRVWPNKGMHAISSFLAVINNALIRANLQK 1620
 Db 1561 HLKAKSSADRFILSGRPMTGDLTRNNKRVWPNKGMHAISSFLAVINNALIRANLQK 1620
 QY 1621 GENPSHYGTAFNHPNLNKOOLSEVALMTTSVDLVASICVIFAMSEVPASVYVFLIOER 1680
 Db 1621 GENPSHYGTAFNHPNLNKOOLSEVALMTTSVDLVASICVIFAMSEVPASVYVFLIOER 1680
 QY 1681 VSKAKHLOFISGVKPVYIWLNSFVMDKCNVYPATLVIIITICQOKSYSTNLPIVAL 1740
 Db 1681 VSKAKHLOFISGVKPVYIWLNSFVMDKCNVYPATLVIIITICQOKSYSTNLPIVAL 1740
 QY 1741 LLLXGKSTPLMPASFEFKIPSTAYVLTSVNLFIGINSVATFVLELTPDKLNIN 1800
 Db 1741 LLLXGKSTPLMPASFEFKIPSTAYVLTSVNLFIGINSVATFVLELTPDKLNIN 1800
 QY 1801 DILKSVFLIPHFCLRGILIDVKNQAMADAFEGENRFPVPSLWDLVGRNLFMAAVEG 1860
 Db 1801 DILKSVFLIPHFCLRGILIDVKNQAMADAFEGENRFPVPSLWDLVGRNLFMAAVEG 1860
 QY 1861 VVFLITVLIOYRFTIRPRVNAKLSPLNDEDEDVREBORILIDGGGNDLEIKELTKI 1920
 Db 1861 VVFLITVLIOYRFTIRPRVNAKLSPLNDEDEDVREBORILIDGGGNDLEIKELTKI 1920
 QY 1921 YRRRRKPAVDRIKCVIPGCEFGILGVNAGKSTFKMLTGDTTYRGAFILNKSIISN 1980
 Db 1921 YRRRRKPAVDRIKCVIPGCEFGILGVNAGKSTFKMLTGDTTYRGAFILNKSIISN 1980
 QY 1981 IHEYHONMGYPOPDATTELLTGREHEFEFALLRGVPEKVGVMATRKILGLVYGEK 2040
 Db 1981 IHEYHONMGYPOPDATTELLTGREHEFEFALLRGVPEKVGVMATRKILGLVYGEK 2040
 QY 2041 YAGNYSGNKRLSTAMALIGPPVFLDEPTTGMDPKARRFLMNCALSYYKBSRVILT 2100
 Db 2041 YAGNYSGNKRLSTAMALIGPPVFLDEPTTGMDPKARRFLMNCALSYYKBSRVILT 2100
 QY 2101 SHSMECEALCTRMALWNGFRCLGSVOHLKNRFGGYTYVIRIAGSNPDLKPYODFEG 2160
 Db 2101 SHSMECEALCTRMALWNGFRCLGSVOHLKNRFGGYTYVIRIAGSNPDLKPYODFEG 2160
 QY 2161 LAFPGSVLEKRRNMLQYOLPSSLSLARIFSILSQRKRHIHIDYSVOSTLDOYVNF 2220
 Db 2161 LAFPGSVLEKRRNMLQYOLPSSLSLARIFSILSQRKRHIHIDYSVOSTLDOYVNF 2220
 QY 2221 AKQSDDDHLKDLSLHKNQTVVDAVLTSLFLODEKVESYV 2261
 Db 2221 AKQSDDDHLKDLSLHKNQTVVDAVLTSLFLODEKVESYV 2261

RESULT 2

US-09-846-456-11

Sequence 11, Application US/09846456

Patent No. US20020146792A1

GENERAL INFORMATION:

APPLICANT: Rosier, Marie

APPLICANT: Prades, Catherine

APPLICANT: Lemoine, Cendrine

APPLICANT: Naudin, Laurent

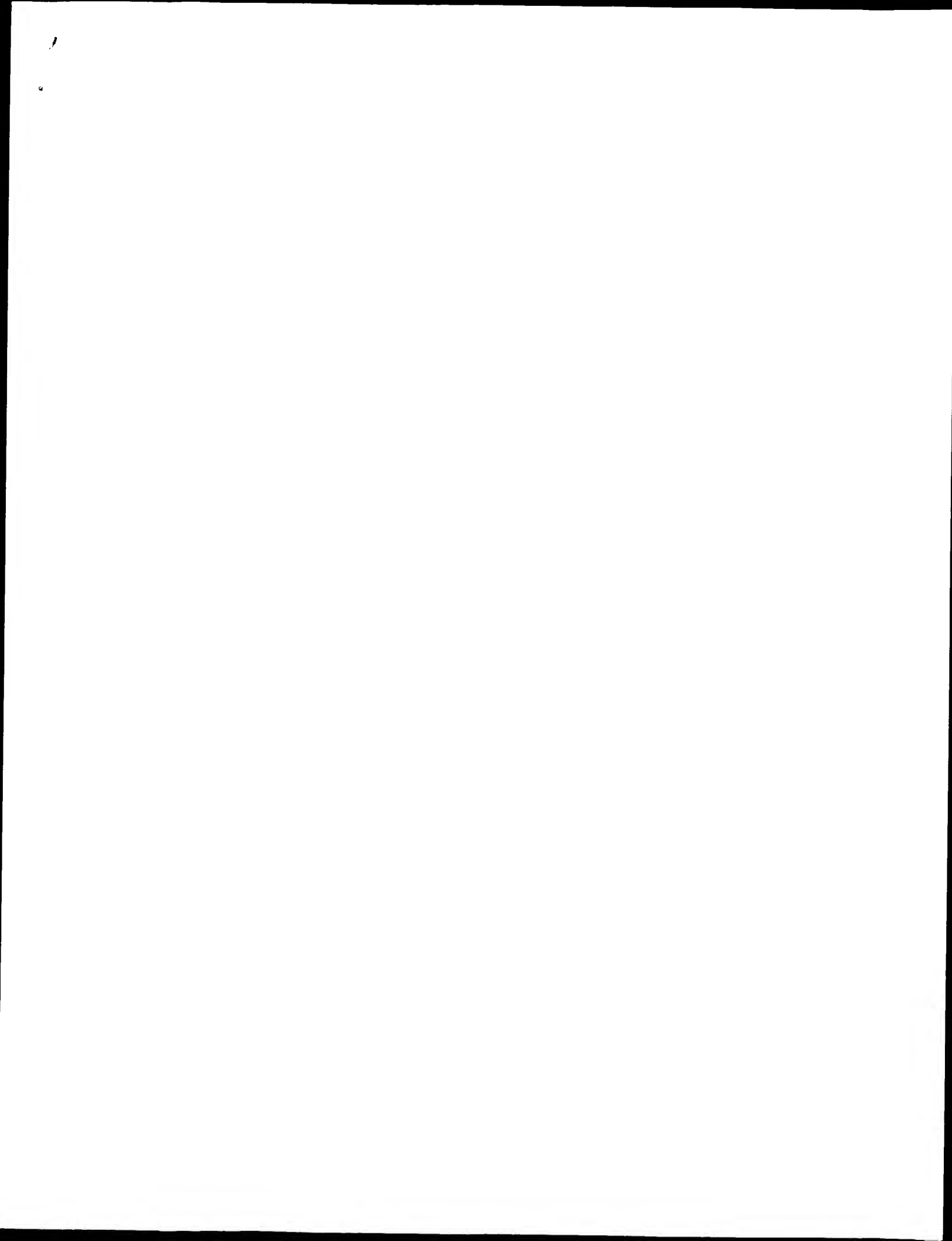
2

APPLICANT: Deneffe, Patrice
 APPLICANT: Duverger, Nicolas
 APPLICANT: Brewer, Brian
 APPLICANT: Remaley, Alan
 APPLICANT: Fojo, Sylvia
 TITLE OF INVENTION: Regulatory Nucleic Acid for the ABC1 Gene, Molecules Modifying It
 TITLE OF INVENTION: Activity and Therapeutic Uses
 FILE REFERENCE: 3806.0505
 CURRENT APPLICATION NUMBER: US/09/846,456
 CURRENT FILING DATE: 2001-05-02
 PRIOR APPLICATION NUMBER: US 60/201,280
 PRIOR FILING DATE: 2000-05-02
 NUMBER OF SEQ ID NOS: 20
 SOFTWARE: PatentIn version 3.0
 SEQ ID NO: 11
 LENGTH: 2261
 TYPE: PRF
 ORGANISM: Homo sapiens
 US-09-846-456-11

Query Match 99.7%; Score 11767; DB 10; Length 2261;
 Best Local Similarity 99.7%; Pred. No. 0;
 Matches 2254; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 MACPOLRLMLKNTLFRRCOTQLLEVAWPLFILLISVLSYSPYEOHCHFPNKA 60
 DB 1 MACPOLRLMLKNTLFRRCOTQLLEVAWPLFILLISVLSYSPYEOHCHFPNKA 60
 QY 61 MPAGTLPMVOGIIICANNPCFRYPTEGAPGVGNFNKSIYARLFSRARLLYSQKDT 120
 DB 61 MPAGTLPMVOGIIICANNPCFRYPTEGAPGVGNFNKSIYARLFSRARLLYSQKDT 120
 QY 121 SKMDMKVRLTLOQIKKSSNKLQDFLVNDFEFGFLVHNLSPKSTYDKMLRAVILH 180
 DB 121 SKMDMKVRLTLOQIKKSSNKLQDFLVNDFEFGFLVHNLSPKSTYDKMLRAVILH 180
 QY 181 KYVLOGYQLHLTSLCNGSKSEEMTOLGDQVSELGIPKELAAARVLRNSNDILKPTL 240
 DB 181 KYVLOGYQLHLTSLCNGSKSEEMTOLGDQVSELGIPKELAAARVLRNSNDILKPTL 240
 QY 241 RTINSTSPPSKELAEATKTLHSLGTLAQELFSMRSWSDMQEVMFLTNVSSSSTQI 300
 DB 241 RTINSTSPPSKELAEATKTLHSLGTLAQELFSMRSWSDMQEVMFLTNVSSSSTQI 300
 QY 301 YQAVSRIVCGHPGGGGLKISLWYEDNNKALFGGNGTEDEAETTYDNTSTPYCNDLAK 360
 DB 301 YQAVSRIVCGHPGGGGLKISLWYEDNNKALFGGNGTEDEAETTYDNTSTPYCNDLAK 360
 QY 361 NLESSPLSRITMKALPLVGLIYTPDTPATROVAEVAEKKTFQELAVFHDLGMMWELS 420
 DB 361 NLESSPLSRITMKALPLVGLIYTPDTPATROVAEVAEKKTFQELAVFHDLGMMWELS 420
 QY 421 PKIWTFMENSQENDLVMLDSDRDNHFWEOQLDGLDWTADIVAFILAKRDEVOSSNGS 480
 DB 421 PKIWTFMENSQENDLVMLDSDRDNHFWEOQLDGLDWTADIVAFILAKRDEVOSSNGS 480
 QY 481 VYTWREAFNETNOAIRTISRMECVNLKLEPIATEVWILINKSMELLDEKKFAAGIYFTG 540
 DB 481 VYTWREAFNETNOAIRTISRMECVNLKLEPIATEVWILINKSMELLDEKKFAAGIYFTG 540
 QY 541 ITPGSIELPHHVKKIKMDIDNVERTNKIKDGYWDPGRADPEEDMKRYVWGFAIYLDVY 600
 DB 541 ITPGSIELPHHVKKIKMDIDNVERTNKIKDGYWDPGRADPEEDMKRYVWGFAIYLDVY 600
 QY 601 EOLATIRVLGEEKTGYVMOOMPVCYVDIPLRVMSRSPLEMTLAWISVAVIIGIY 660
 DB 601 EOLATIRVLGEEKTGYVMOOMPVCYVDIPLRVMSRSPLEMTLAWISVAVIIGIY 660
 QY 661 YEKEARLKETKIRINGLSDNLSWFSISLPLVLSAGLLVILKGNLLPYSDPSVVEV 720
 DB 661 YEKEARLKETKIRINGLSDNLSWFSISLPLVLSAGLLVILKGNLLPYSDPSVVEV 720
 QY 721 FLVSFAVYVTTIQCFLISTLSRANLAAACGGIYFTLYLPYVLCVAMQDYGFTLKIFAS 780

DB 721 FLVSFAVYVTTIQCFLISTLSRANLAAACGGIYFTLYLPYVLCVAMQDYGFTLKIFAS 780
 QY 781 LLSVAFGCGEYFALFEEOGIGVQNDNLESPEVEDGNLTTSISMALFDTLGYVMTW 840
 DB 781 LLSVAFGCGEYFALFEEOGIGVQNDNLESPEVEDGNLTTSISMALFDTLGYVMTW 840
 QY 841 YIEAVFPGQYIIPRPWFPCPKSYWPEEEDSKSPNSOKRISLCEIMBEEPHLKIGVS 900
 DB 841 YIEAVFPGQYIIPRPWFPCPKSYWPEEEDSKSPNSOKRISLCEIMBEEPHLKIGVS 900
 QY 901 IONLVKVRDGMKVAVQALNFEQOITSPFLHNGAGKTTMSILTLGPPPTSGTYIL 960
 DB 901 IONLVKVRDGMKVAVQALNFEQOITSPFLHNGAGKTTMSILTLGPPPTSGTYIL 960
 QY 961 GKDIRSEMTIRONLGYCPOHNVLFDMLTVEEHIWYARLKGISEKHVAKMEOMALDVG 1020
 DB 961 GKDIRSEMTIRONLGYCPOHNVLFDMLTVEEHIWYARLKGISEKHVAKMEOMALDVG 1020
 QY 1021 LPSSKLSKTSQLSGGMQRKLSVALAFVGSKYIIDEPTAGVDPYRGRIMWELLKYRQ 1080
 DB 1021 LPSSKLSKTSQLSGGMQRKLSVALAFVGSKYIIDEPTAGVDPYRGRIMWELLKYRQ 1080
 QY 1081 GRTIILSTHMDADVLDGRIAIISHGKLCVGSLSFLKNOIGTYLTVKKDVESSL 1140
 DB 1081 GRTIILSTHMDADVLDGRIAIISHGKLCVGSLSFLKNOIGTYLTVKKDVESSL 1140
 QY 1141 SCRNSSTVSTLKEDSVSSSDAGLSHESDTLIDVSAISNLIRKHSEARLYEDI 1200
 DB 1141 SCRNSSTVSTLKEDSVSSSDAGLSHESDTLIDVSAISNLIRKHSEARLYEDI 1200
 QY 1201 GHELTYYVPEAKRGAVELFHEIDRLSDLGISYGISSTTLEELFLKAAEESGDAE 1260
 DB 1201 GHELTYYVPEAKRGAVELFHEIDRLSDLGISYGISSTTLEELFLKAAEESGDAE 1260
 QY 1261 TSDGTLPARNRKRAFGDKOSCLPFTEDDADPNDSDIDPESRETDLSCMDGKSGYQV 1320
 DB 1261 TSDGTLPARNRKRAFGDKOSCLPFTEDDADPNDSDIDPESRETDLSCMDGKSGYQV 1320
 QY 1321 GKKLTQOQVALLMKRLIARSRKGFPAQIVLPAVFCIALVSLVPPGKPSLEQ 1380
 DB 1321 GKKLTQOQVALLMKRLIARSRKGFPAQIVLPAVFCIALVSLVPPGKPSLEQ 1380
 QY 1381 PMYNQOYTFVSNDAPEDTGTELLNALTKDPGFCRMBENPIPDPTCQAGEEMTAP 1440
 DB 1381 PMYNQOYTFVSNDAPEDTGTELLNALTKDPGFCRMBENPIPDPTCQAGEEMTAP 1440
 QY 1441 VPOTIMDLFONQNTMNPSPACQSSDKIKMLPVCPPGAGGLPPQKONTADILQDL 1500
 DB 1441 VPOTIMDLFONQNTMNPSPACQSSDKIKMLPVCPPGAGGLPPQKONTADILQDL 1500
 QY 1501 TGRNISDYLYKTYVQITAKSLKNKTWNEFRYGGSLSVNTQALPPOEVENDATKQKK 1560
 DB 1501 TGRNISDYLYKTYVQITAKSLKNKTWNEFRYGGSLSVNTQALPPOEVENDATKQKK 1560
 QY 1561 HLKLAQSSADRFSLSLGTFMTGDTNRNNVYKWFNNKGMAHAISSFLYINNAITRANLQK 1620
 DB 1561 HLKLAQSSADRFSLSLGTFMTGDTNRNNVYKWFNNKGMAHAISSFLYINNAITRANLQK 1620
 QY 1621 GENPSHYGTTAENHPLNLTKOOLSEVALMTTSVDVLSICYFAMSFPASFPVFLIOER 1680
 DB 1621 GENPSHYGTTAENHPLNLTKOOLSEVALMTTSVDVLSICYFAMSFPASFPVFLIOER 1680
 QY 1681 VSKAKHLOFISGVKPYIYLSNFWDMCNVYPALVLIIFICFOOKSYVSTNLPVAL 1740
 DB 1681 VSKAKHLOFISGVKPYIYLSNFWDMCNVYPALVLIIFICFOOKSYVSTNLPVAL 1740
 QY 1741 LLLYGSITPLMYPASVFKIPSTAYVVLVSUNLFGINSVATFVLELFTDNKLNIN 1800
 DB 1741 LLLYGSITPLMYPASVFKIPSTAYVVLVSUNLFGINSVATFVLELFTDNKLNIN 1800
 QY 1801 DILKSYFLFPHFCUGRGLIDWYKQAMADALERGGERFVPSLWDVGRULFAMAEG 1860



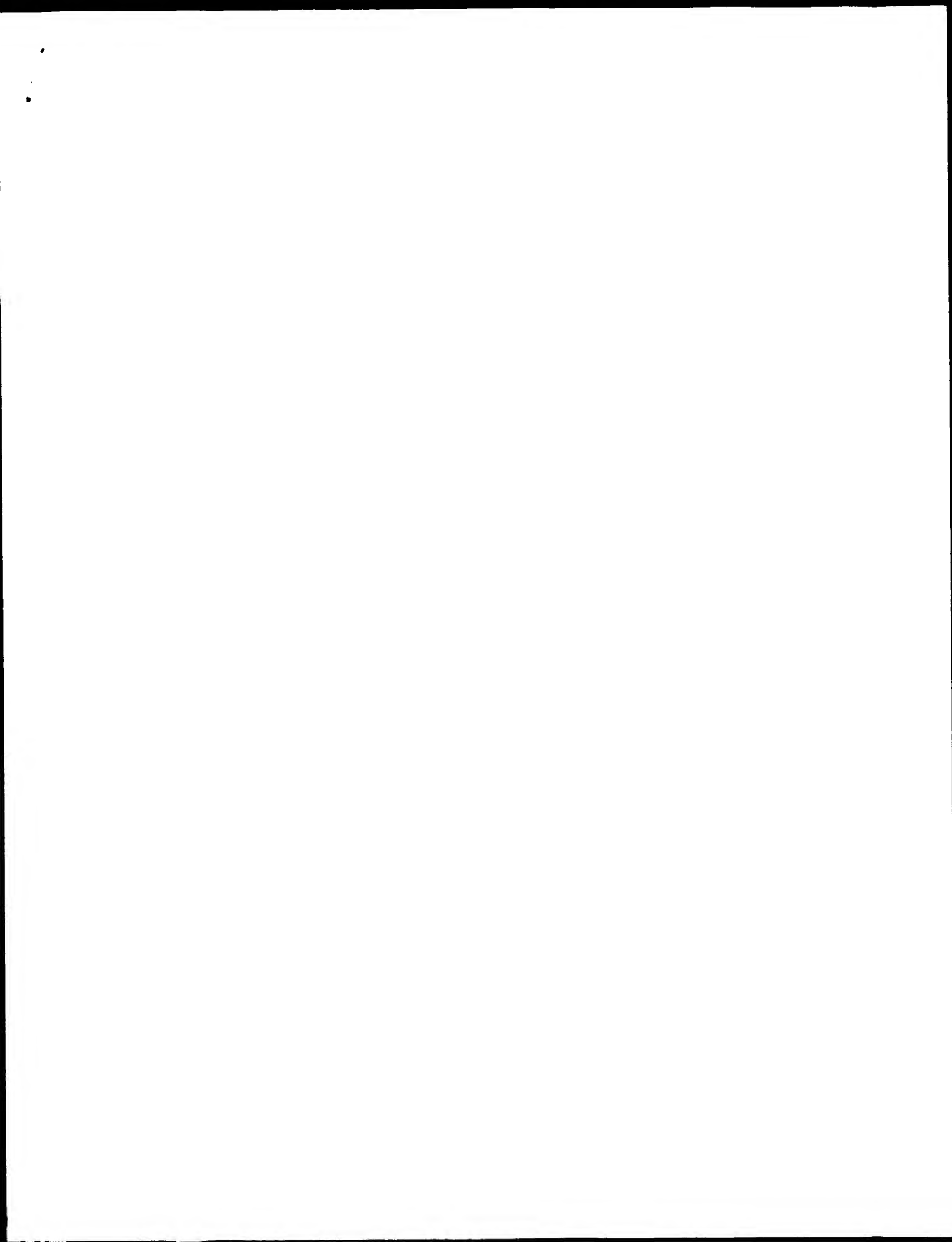
Db 1801 DILKSVFLIFPHFGLGSLIDMVKNOAMADALBERFGENRFSPLSMDIVGRNLFAMAVEG 1860
Qy 1861 VEFELITVLIOYREFIIRPPVNAKLSPLNDEDEVDREORILIDGGONDILEIKELTKI 1920
Db 1861 VEFELITVLIOYREFIIRPPVNAKLSPLNDEDEVDREORILIDGGONDILEIKELTKI 1920
Qy 1921 YRRKKRPVADRICVGIIPGBCFGLGVNAGKSGSTKMLTGTVTYRGDAFLKNSILSN 1980
Db 1921 YRRKKRPVADRICVGIIPGBCFGLGVNAGKSGSTKMLTGTVTYRGDAFLKNSILSN 1980
Qy 1981 IHEVHOMVGCPOFDATTELLTGREHVEFALLRGVPEKGVGEMALTKGLVYKGEK 2040
Db 1981 IHEVHOMVGCPOFDATTELLTGREHVEFALLRGVPEKGVGEMALTKGLVYKGEK 2040
Qy 2041 YAGNYSNGNRKRLSTAMALIGPVPVLEDEPTGMDPKARFLMNCALSVYKGRSVLT 2100
Db 2041 YAGNYSNGNRKRLSTAMALIGPVPVLEDEPTGMDPKARFLMNCALSVYKGRSVLT 2100
Qy 2101 SHSMECEALCTMAALVNGRFGCLSGVOHLKRRFGDGYTIIVRIAGSNEDLKPVODFG 2160
Db 2101 SHSMECEALCTMAALVNGRFGCLSGVOHLKRRFGDGYTIIVRIAGSNEDLKPVODFG 2160
Qy 2161 LAPPGSVKEKHNNMLOLOPSSLSLARFEFSLSSKRLHEDYSVOTITDOYFVNF 2220
Db 2161 LAPPGSVKEKHNNMLOLOPSSLSLARFEFSLSSKRLHEDYSVOTITDOYFVNF 2220
Qy 2221 AKDOSDDHLKDLTLKNNQTVVAVLTSFLQDEKYESYV 2261
Db 2221 AKDOSDDHLKDLTLKNNQTVVAVLTSFLQDEKYESYV 2261

RESULT 3
US-09-595-542-9
Sequence 9, Application US/0995542
Patent No. US20020187647A1
GENERAL INFORMATION:
APPLICANT: Shutter, John
APPLICANT: Ollas, Laetia
TITLE OF INVENTION: ATP-Binding Cassette Transporter-Like Molecules and
FILE REFERENCE: 00-658-A
CURRENT APPLICATION NUMBER: US-09/995,542
CURRENT FILING DATE: 2000-11-28
PRIOR APPLICATION NUMBER: 60/253,520
PRIOR FILING DATE: 2000-11-28
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 9
LENGTH: 2201
TYPE: PRT
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: UNSURE
LOCATION: (115)
OTHER INFORMATION: amino acid at this position is unknown
US-09-995-542-9

Query Match 92.6%; Score 10922; DB 10; Length 2201;
Best Local Similarity 94.9%; Pred. No. 0;
Matches 2089; Conservative 55; Mismatches 57; Indels 0; Gaps 0;

Qy 61 MPASAGILPWVOGIIICANNPCFRYPPEGEAPGVVFNKSIARLFSARLLILYKQKDT 120
Db 1 MPASAGILPWVOGIIICANNPCFRYPPEGEAPGVVFNKSIARLFSARLLILYKQKDT 60
Qy 121 SMDMRVLTLOQIKSSNLTLOPFIYDNEFFSGFLVHNLSLPSTYDKMLRAVYIDK 180
Db 61 SMDMRVLTLOQIKSSNLTLOPFIYDNEFFSGFLVHNLSLPSTYDKMLRAVYIDK 180
Qy 181 KVFLOGVQLHLTSLGSKSEMIQGLDQVSESLGKLPKELAAARVLRNSMDILKPII 240
Db 121 KVFLOGVQLHLTSLGSKSEMIQGLDQVSESLGKLPKELAAARVLRNSMDILKPII 180

Qy 241 RTLNSTSPSPSKELAEATKTLHSLGLTLOELFMSRMSKMDKROEMFLTNVSSSSSTOI 300
Db 181 RTLNSTSPSPSKELAEATKTLHSLGLTLOELFMSRMSKMDKROEMFLTNVSSSSSTOI 240
Qy 301 YQAVSRIVCGHPEGGLIKISLWYEDNNYKALFGNGTEDEAETFYDNSTTPCNDLMK 360
Db 241 YQAVSRIVCGHPEGGLIKISLWYEDNNYKALFGNGTEDEAETFYDNSTTPCNDLMK 300
Qy 361 NLESSPRLIITKALKPLLYGKILYTPPTPATROYMAEVNKTFOELAVFHDLBGMWELS 420
Db 301 NLESSPRLIITKALKPLLYGKILYTPPTPATROYMAEVNKTFOELAVFHDLBGMWELS 360
Qy 421 PKIWFEMNSQEMDVRLNLSRDNDHFWEODLDMPTAADIYAFLAKHEDVOSSNGS 480
Db 361 PKIWFEMNSQEMDVRLNLSRDNDHFWEODLDMPTAADIYAFLAKHEDVOSSNGS 420
Qy 481 VYTRAEAFNETNOALITISREMECVNLKLEPIATEVWLINKSMELDERKFMAGIVFTG 540
Db 421 VYTRAEAFNETNOALITISREMECVNLKLEPIATEVWLINKSMELDERKFMAGIVFTG 480
Qy 541 TPGSIELPHHKYKIRMDIDNVERTNKIKDGYNDPGPRADPFEDMRVWGGFAYLDDYV 600
Db 481 TPGSIELPHHKYKIRMDIDNVERTNKIKDGYNDPGPRADPFEDMRVWGGFAYLDDYV 540
Qy 601 EQAIIIVLTGTEKKTGYVMQMPYCYVDJFELRMSKMPFLPTLAMYSVAYIIRGIV 660
Db 541 EQAIIIVLTGTEKKTGYVMQMPYCYVDJFELRMSKMPFLPTLAMYSVAYIIRGIV 600
Qy 661 YEKEARKETMRIMGDLSILMFSWFTSSKIPLLVSAGLVYILKGNLIPYSDPVFV 720
Db 601 YEKEARKETMRIMGDLSILMFSWFTSSKIPLLVSAGLVYILKGNLIPYSDPVFV 660
Qy 721 FLTSFAVNTILOCFLISLFLSRANLAACGGITFTLYLPRYLGVANODYGFGLTKIFAS 780
Db 661 FLTSFAVNTILOCFLISLFLSRANLAACGGITFTLYLPRYLGVANODYGFGLTKIFAS 720
Qy 781 LLSPAVFGGCEYFALFEEOGIGVOMDNLFPSPVEEDPFNTITSIMKLEPTFLYGYMTV 840
Db 721 LLSPAVFGGCEYFALFEEOGIGVOMDNLFPSPVEEDPFNTITSIMKLEPTFLYGYMTV 780
Qy 841 YLEAVFGQYIGIPRWYPPCTKSTWFGSESEKSHPSGNOKRSEICMEEPFLKLGVS 900
Db 781 YLEAVFGQYIGIPRWYPPCTKSTWFGSESEKSHPSGNOKRSEICMEEPFLKLGVS 840
Qy 901 IONLVKVRDGAKVAVDGLALNEFGQITTSFLGNHAGKTTMGLSLPPTSGTAYIL 960
Db 841 IONLVKVRDGAKVAVDGLALNEFGQITTSFLGNHAGKTTMGLSLPPTSGTAYIL 900
Qy 961 GKDISEMSTIRONLGYCPOHNVLEFDMLTVEHIMFYARLKGSEKHVAKMEQMALDVG 1020
Db 901 GKDISEMSTIRONLGYCPOHNVLEFDMLTVEHIMFYARLKGSEKHVAKMEQMALDVG 960
Qy 1021 LPSSKLAKTSQSLSGMOKRKSVALAFVGGSKVYILDEPTAGVDPYSRGIWELDKYRQ 1080
Db 961 LPSSKLAKTSQSLSGMOKRKSVALAFVGGSKVYILDEPTAGVDPYSRGIWELDKYRQ 1020
Qy 1081 GRTIILSTHMDADVLDRLAIISHGKLCVGSLSPLKNOLOGVYTLVKKOVESLS 1140
Db 1021 GRTIILSTHMDADVLDRLAIISHGKLCVGSLSPLKNOLOGVYTLVKKOVESLS 1080
Qy 1141 SCRNSSSTVSYLKEDSVSSSDAGLSDESDTLTIDVSAISNLIRKHVSEARLEDI 1200
Db 1081 SCRNSSSTVSYLKEDSVSSSDAGLSDESDTLTIDVSAISNLIRKHVSEARLEDI 1140
Qy 1201 GHELTYYVPEAKAGAVELFHEIDRLSDLGISSTGISTTLEETFLKVAESGVDAE 1260
Db 1141 GHELTYYVPEAKAGAVELFHEIDRLSDLGISSTGISTTLEETFLKVAESGVDAE 1200
Qy 1261 TSDGTLPARNRRAFGDQSCRPTEEDDADPNDSDIDPSRPTDLISGNDGSGSYOVK 1320
Db 1201 TSDGTLPARNRRAFGDQSCRPTEEDDADPNDSDIDPSRPTDLISGNDGSGSYOVK 1260
Qy 1321 GWKLTQGGFVALMKRLLIARRSRKGFPAQIVLPVAVFCIALVSLVPPGKXPSLEIQ 1380



GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 7, 2003, 08:41:19 ; Search time 41 Seconds
(without alignments)
2287.268 Million cell updates/sec

Title: US-09-595-526c-2

Perfect score: 11797
Sequence: 1 MACWPQLRLTLWKNTLFRRR.....VDVAVLTSFLDPEFKESYV 2261

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	11770	99.8	2261	1	ABCI_HUMAN
2	11256	95.4	2261	1	ABCI_MOUSE
3	5870.5	49.8	2273	1	ABCR_HUMAN
4	4230.5	35.7	2436	1	ABCR2_HUMAN
5	4093.5	34.9	2434	1	ABCR2_MOUSE
6	2642.5	22.4	1704	1	ABCR3_HUMAN
7	1538.5	13.0	1704	1	CDP7_CAERL
8	411	3.5	330	1	DRRA_STRE
9	382.5	3.2	343	1	NOD1_RHISN
10	366	3.1	304	1	NOD1_RHIS3
11	347	2.9	308	1	YADG_ECOLI
12	344.5	2.9	340	1	NOD1_RHITO
13	343.5	2.9	347	1	NOD1_RHICA
14	339.5	2.9	1280	1	MDR1_HUMAN
15	331.5	2.8	355	1	MDR1_RHIME
16	329.5	2.8	578	1	YBHF_ECOLI
17	327.5	2.8	1281	1	MDR3_CRIGR
18	327	2.8	894	1	YHTR_ECOLI
19	326	2.8	1276	1	MDR3_MOUSE
20	325.5	2.8	308	1	NOSE_PSEST
21	324.5	2.8	1276	1	MDR2_MOUSE
22	322.5	2.7	306	1	MDR1_BRAVA
23	321	2.7	1276	1	MDR1_CRIGR
24	317	2.7	1362	1	PMO1_SCHPO
25	316	2.7	1276	1	MDR2_CRIGR
26	315.5	2.7	311	1	NOD1_RHILV
27	315	2.7	1278	1	MDR1_RAT
28	314.5	2.7	1277	1	MDR1_RAT
29	313	2.7	354	1	Y415_SYNY3
30	310.5	2.6	1276	1	MDR1_MOUSE
31	304.5	2.6	1294	1	YOH5_YEAST
32	300	2.5	381	1	OPAB_BACSU
33	298.5	2.5	262	1	YA23_METUA

34	297	2.5	380	1	OPCA_BACSU
35	296	2.5	335	1	V719_ANASP
36	293.5	2.5	1321	1	AB11_HUMAN
37	292.5	2.5	305	1	YHCH_BACSU
38	287.5	2.4	274	1	V179_MYCPN
39	286.5	2.4	274	1	V179_MYCCE
40	285.5	2.4	1279	1	MDR3_HUMAN
41	284	2.4	306	1	BCRA_BACLI
42	284	2.4	1321	1	MDR1_CAERL
43	284	2.4	1336	1	MAM1_SCHPO
44	281.5	2.4	308	1	YEHX_ECOLI
45	281.5	2.4	343	1	ABC_ECOLI

ALIGNMENTS

RESULT 1
ABCI_HUMAN STANDARD: PRT: 2261 AA.
AC 095477: 09UN08: 09UN07: 09UN06: 09NOV4: 09UN09: 096T85: 096556:
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ATP-binding cassette, sub-family A, member 1 (ATP-binding cassette transporter 1) (ATP-binding cassette 1) (ABC-1) (Cholesterol efflux regulatory protein).
DE ABCA1 OR ABC1 OR CERP.
GN Homo sapiens (Human)
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_Taxid:9606;
RN [1]
RP MEDLINE=2034509; PubMed=10884428;
RX Sanamaria-Rojas S., Peterson K.M., Knapper C.L., Olu Y., Freeman L.A., Cheng J.-F., Osorio J., Remley A.T., Yang X.-P., Haendenschild C.C., Prades C., Chimini G., Blackton E.F., Francois T.L., Duverger N., Rubin E.M., Koster M., Deneffe P., Fredrickson D.S., Brewer H.B. Jr.;
RA "Complete genomic sequence of the human ABCA1 gene: analysis of the human and mouse ATP-binding cassette A promoter.";
RT Proc. Natl. Acad. Sci. U.S.A. 97:7987-7992(2000). -544.
RL [2]
RN [2]
RP TISSUE=Skin;
RA Schwartz K., Lawn R.M., Wade D.P.;
RT "ABCA1 gene expression and apoA-I-mediated cholesterol efflux are regulated by LXR.";
RT Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP MEDLINE=9191454; PubMed=11352567;
RX Olu Y., Cavelier L., Chiu S., Yang X., Rubin E., Cheng J.-F.;
RA "Human and mouse ABCA1 comparative sequencing and transgenesis studies: revealing novel regulatory sequences.";
RT Genomics 73:66-76(2001).
RN [4]
RP MEDLINE=9191454; PubMed=10092505;
RX Langmann T., Kattnicken J., Reil M., Liebisch G., Luciani M.F., Chimini G., Klinken W.E., Schmitz G.;
RA "Molecular cloning of the human ATP-binding cassette transporter 1 (ABCA1): evidence for sterol-dependent regulation in macrophages.";
RT Biochem. Biophys. Res. Commun. 257:29-33(1999).
RN [6]

RP SEQUENCE OF 21-2261 FROM N.A.
 RA MEDLINE=99364413; PubMed=10431238;
 RA Rust S., Rosier M., Funke H., Reel J., Amoura Z., Pilette J.-C.,
 RA Delzenne J.-F., Brewer H.B., Duvenger N., Denelle P., Assmann G.,
 RT "Tangier disease is caused by mutations in the gene encoding
 RT ATP-binding cassette transporter 1.",
 RL Nat. Genet. 22:352-355(1999).
 RN [17]
 RP VARIANTS FHA THR-1091 AND 1893-GLU-ASP-1894 DEL.
 RX MEDLINE=20001430; PubMed=10533863;
 RA Marcell M., Brooks-Wilson A., Clee S.M., Roomp K., Zhang L.-H., Yu L.,
 RA Collins J.A., van Dam M., Molhuizen H.O.F., Loubser O.,
 RA Ouellette B.F.F., Senses C.W., Fichter K., Mott S., Denis M.,
 RA Bouchet B., Pimstone S., Genest J. Jr., Kastelein J.J.P., Hayden M.R.,
 RT "Mutations in the ABC1 gene in familial HDL deficiency with defective
 RT cholesterol efflux.",
 RL Lancet 354:1341-1346(1999).
 RN [8]
 RP VARIANTS TD ARG-597 AND ARG-1477, AND VARIANT FHA LEU-693 DEL.
 RX MEDLINE=99364411; PubMed=10431236;
 RA Brooks-Wilson A., Marcell M., Clee S.M., Zhang L.-H., Roomp K.,
 RA van Dam M., Yu L., Brewer C., Collins J.A., Molhuizen H.O.F.,
 RA Loubser O., Ouellette B.F.F., Fichter K., Ashbourne-Excoffon K.J.D.,
 RA Senses C.W., Scherer S., Mott S., Denis M., Martindale D.,
 RA Frohlich J., Morgan K., Koop B., Pimstone S., Kastelein J.J.P.,
 RA Hayden M.R.,
 RT "Mutations in ABC1 in Tangier disease and familial high-density
 RT lipoprotein deficiency.",
 RL Nat. Genet. 22:336-345(1999).
 RN [9]
 RP VARIANTS TD SER-590; SER-935 AND VAL-937, AND VARIANTS ALA-399 AND
 RX MET-883.
 RA MEDLINE=99364412; PubMed=10431237;
 RA Bodzioch M., Orso E., Klucken J., Langmann T., Boettcher A.,
 RA Diedrich W., Drobnik W., Barlage S., Buechler A.,
 RA Porcsch-Geszczewicz M., Kaminski W.E., Hehmann H.W., Oette K.,
 RA Rothe G., Aslanidis C., Lackner K.J., Schmitz G.,
 RT "The gene encoding ATP-binding cassette transporter 1 is mutated in
 RT Tangier disease.",
 RL Nat. Genet. 22:347-351(1999).
 RN [10]
 RP VARIANTS TD ILE-929; ARG-597 AND ARG-1477, AND VARIANTS FHA LEU-693
 RX DEL; THR-1091; 1893-GLU-ASP-1894 DEL AND LEU-2150.
 RA MEDLINE=20540002; PubMed=11086027;
 RA Clee S.M., Kastelein J.J.P., van Dam M., Marcell M., Roomp K.,
 RA Zwarts K.Y., Collins J.A., Roelants R., Tamasa N., Stulic T.,
 RA Suda T., Ceska R., Boucher B., Rondeau C., Desoulhac C.,
 RA Brooks-Wilson A., Molhuizen H.O.F., Frohlich J., Genest J. Jr.,
 RA Hayden M.R.,
 RT "Age and residual cholesterol efflux affect HDL cholesterol levels and
 RT coronary artery disease in ABCA1 heterozygotes.",
 RL J. Clin. Invest. 106:1263-1270(2000).
 RN [11]
 RP VARIANTS TD ASN-1289 AND HIS-1800.
 RX MEDLINE=20171564; PubMed=10706591;
 RA Brousseau M.E., Schaefer E.J., Dupuis J., Eustace B.,
 RA Van Berdevegh P., Goldkamp A.L., Thurston L.M., Fitzgerald M.G.,
 RA Yasek-Mckenna D., O'Neill G., Eberhart G.P., Weifenhach B.,
 RA Ordovas J.M., Freeman M.W., Brown R.H. Jr., Gu J.Z.,
 RT "Novel mutations in the gene encoding ATP-binding cassette 1 in four
 RT Tangier disease kindreds.",
 RL J. Lipid Res. 41:433-441(2000).
 RN [12]
 RP VARIANT TD ASP-1046, VARIANT FHA CYS-230, AND VARIANTS LYS-219;
 RX ILE-825; MET-883 AND LYS-1587.
 RA MEDLINE=20396633; PubMed=10938021;
 RA Wang J., Burnett J.R., Near S., Young K., Ziman B., Hanley A.J.G.,
 RA Connelly P.W., Harris S.B., Hegele R.A.,
 RT "Common and rare ABCA1 variants affecting plasma HDL cholesterol.",
 RL Arterioscler. Thromb. Vasc. Biol. 20:1983-1989(2000).
 RN [13]
 RP VARIANT TD TRP-587, AND VARIANT LEU-2168.
 RX MEDLINE=21157002; PubMed=11257260;
 RA Bertolini S., Plasciotta L., Seri M., Cusano R., Cantafora A.,
 RA Calabresi L., Franceschini G., Ravazzolo R., Calandra S.,
 RT "A point mutation in ABC1 gene in a patient with severe premature
 RT coronary heart disease and mild clinical phenotype of Tangier
 RT disease.",
 RL Atherosclerosis 154:599-605(2001).
 RN [14]
 RP VARIANTS LYS-219; MET-883 AND ASP-1172.
 RX MEDLINE=21157003; PubMed=11257261;
 RA Brousseau M.E., Bodzioch M., Schaefer E.J., Goldkamp A.L., Kielar D.,
 RA Probst M., Ordovas J.M., Aslanidis C., Lackner K.J.,
 RA Bloomfield Robbins H., Collins D., Robins S.J., Wilson P.W.F.,
 RA Schmitz G.,
 RT "Common variants in the gene encoding ATP-binding cassette transporter
 RT 1 in men with low HDL cholesterol levels and coronary heart disease.",
 RL Atherosclerosis 154:607-611(2001).
 RN [15]
 RP VARIANT TD LEU-1506.
 RX MEDLINE=21369429; PubMed=11476961;
 RA Lapicka-Bodzioch K., Bodzioch M., Kruehl M., Kielar D., Probst M.,
 RA Klec B., Andrikovics H., Boettcher A., Hubacek J., Aslanidis C.,
 RA Suttorp N., Schmitz G.,
 RT "Homogenous assay based on 52 primer sets to scan for mutations of
 RT the ABCA1 gene and its application in genetic analysis of a new
 RT patient with familial high-density lipoprotein deficiency syndrome.",
 RL Biochim. Biophys. Acta 1537:42-48(2001).
 RN [16]
 RP VARIANTS TD ASN-1289 AND TRP-2081, AND VARIANT LYS-219.
 RX MEDLINE=21369433; PubMed=11476965;
 RA Huang W., Moriyama K., Koga T., Hua H., Ageta M., Kawabata S.,
 RA Matsutani K., Yamura T., Eto T., Kawamura M., Teramoto T., Sasaki J.,
 RT "Novel mutations in ABCA1 gene in Japanese patients with Tangier
 RT disease and familial high density lipoprotein deficiency with
 RT coronary heart disease.",
 RL Biochim. Biophys. Acta 1537:71-78(2001).
 RN [17]
 RP VARIANTS LYS-219; ALA-399; MET-771; PRO-774; ASN-776; ILE-825;
 RX MET-883; ASP-1172; LYS-1587 AND CYS-1731.
 RA MEDLINE=21138379; PubMed=11238261;
 RA Clee S.M., Zwinderman A.H., Engert J.C., Zwarts K.Y.,
 RA Molhuizen H.O.F., Roomp K., Jukema J.W., van Wijland M., van Dam M.,
 RA Hudson T.J., Brooks-Wilson A., Genest J. Jr., Kastelein J.J.P.,
 RA Hayden M.R.,
 RT "Common genetic variation in ABCA1 is associated with altered
 RT lipoprotein levels and a modified risk for coronary artery disease.",
 RL Circulation 103:1198-1205(2001).
 RN [18]
 RP VARIANT TD THR-255, AND VARIANT ATHEROSCLEROSIS ASP-1611.
 RX MEDLINE=21645894; PubMed=11785958;
 RA Nishida Y., Hirano K., Tsukamoto K., Nagano M., Ikegami C., Roomp K.,
 RA Ishihara M., Sakane N., Zhang Z., Tsujii K., Matsuyama A., Ohama T.,
 RA Matsuura F., Ishigami M., Sakai N., Hiraoka H., Hattori H.,
 RA Wellington C., Yoshida Y., Misugi S., Hayden M.R., Egashira T.,
 RA Yamashita S., Matsuzawa Y.,
 RT "Expression and functional analyses of novel mutations of ATP-binding
 RT cassette transporter-1 in Japanese patients with high-density
 RT lipoprotein deficiency.",
 RL Biochem. Biophys. Res. Commun. 290:713-721(2002).
 CC -1- FUNCTION: CAMP-DEPENDENT AND SULFONYLUREA-SENSITIVE ANION
 CC TRANSPORT. KEY GATEKEEPER INFLUENCING INTRACELLULAR CHOLESTEROL
 CC TRANSPORT.
 CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED, BUT MOST ABUNDANT IN
 CC MACROPHAGES.
 CC -1- DOMAIN: MULTIFUNCTIONAL, POLYPEPTIDE WITH TWO HOMOLOGOUS HALVES,
 CC EACH CONTAINING AN HYDROPHOBIC MEMBRANE-ANCHORING DOMAIN AND AN
 CC ATP BINDING CASSETTE (ABC) DOMAIN.
 CC -1- DISEASE: DEFECTS IN ABCA1 ARE A CAUSE OF HIGH DENSITY LIPOPROTEIN
 CC DEFICIENCY TYPE I (HDL1), ALSO KNOWN AS TANGIER DISEASE (TD). TD
 CC IS A RECESSIVE DISORDER CHARACTERIZED BY ABSENCE OF HIGH DENSITY
 CC LIPOPROTEIN (HDL) CHOLESTEROL FROM PLASMA, HEPATOSPLENOMEGALY,
 CC PERIPHERAL NEUROPATHY, AND FREQUENTLY PREMATURE CORONARY ARTERY
 CC DISEASE (CAD).
 CC -1- DISEASE: Defects in ABCA1 are a cause of high density lipoprotein

Query Match 99.8%; Score 11770; DB 1; Length 2261;
 Best Local Similarity 99.7%; Pred. No. 0;
 Matches 2255; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 MACPOLRLILKNNLFRFRROTCQLLEVAWPLFJLLIISVRLSYPEQEHCHFPKKA 60
 DB 1 MACPOLRLILKNNLFRFRROTCQLLEVAWPLFJLLIISVRLSYPEQEHCHFPKKA 60
 QY 61 MPAGTLPWVOGIICNANNPCFRYPTEPGAEPGVGNFNKSIYARLFSDARLLILSQKDT 120
 DB 61 MPAGTLPWVOGIICNANNPCFRYPTEPGAEPGVGNFNKSIYARLFSDARLLILSQKDT 120
 QY 121 SMKDMRYLRTLOOIKKSSSNLKLODFLVNTEFFSGFLHNLSLPSKSTYDKMLRAVILH 180
 DB 121 SMKDMRYLRTLOOIKKSSSNLKLODFLVNTEFFSGFLHNLSLPSKSTYDKMLRAVILH 180
 QY 181 KVFLOGYOLHLSLCSNKSSEMIQLODOEVSSELCGLPKKXLAABRVLRSMNDILKPTL 240
 DB 181 KVFLOGYOLHLSLCSNKSSEMIQLODOEVSSELCGLPKKXLAABRVLRSMNDILKPTL 240
 QY 241 RFLNSTSPFSKELAEATKTLHSLGTLAOLFPMSRMSMDROEVMFLTNVSSSSSTOI 300
 DB 241 RFLNSTSPFSKELAEATKTLHSLGTLAOLFPMSRMSMDROEVMFLTNVSSSSSTOI 300
 QY 301 YQAVRIYCGHPGEGGLKIKSLNMYEDNNYKALFGNGTEDEAETFYDNSTTPYCNLDM 360
 DB 301 YQAVRIYCGHPGEGGLKIKSLNMYEDNNYKALFGNGTEDEAETFYDNSTTPYCNLDM 360
 QY 361 NDESSPLSHIWKALKPLLVGKILYTPPTPATROYMAEVNKTROELAVFHLEMEELS 420
 DB 361 NDESSPLSHIWKALKPLLVGKILYTPPTPATROYMAEVNKTROELAVFHLEMEELS 420
 QY 421 PKIWTMENSOAMDLYRMLDSDRDHFWEOQLDIDMTAODIYAFIAKHEDEVOSSNGS 480
 DB 421 PKIWTMENSOAMDLYRMLDSDRDHFWEOQLDIDMTAODIYAFIAKHEDEVOSSNGS 480
 QY 481 VYTWREAFNETNOAIRTISRMECVNLKLEPIATEVWILINKSMELDEKFMAGIYVTG 540
 DB 481 VYTWREAFNETNOAIRTISRMECVNLKLEPIATEVWILINKSMELDEKFMAGIYVTG 540
 QY 541 ITTGSTELPHHYKRIKIMDIDNVERTNKIKDGVWDGPRADPEEDMRYVWGFAVLQDV 600
 DB 541 ITTGSTELPHHYKRIKIMDIDNVERTNKIKDGVWDGPRADPEEDMRYVWGFAVLQDV 600
 QY 601 EOALIRVLTGTEKKTGYMOQMPYPCYVDIILRVNSRSMPLFMTLAWIYSAVILIKGI 660
 DB 601 EOALIRVLTGTEKKTGYMOQMPYPCYVDIILRVNSRSMPLFMTLAWIYSAVILIKGI 660
 QY 661 YEKARLKEIMRIMGLDNLISLFWISLILPLVNSAGLLVILKIGNLLPYSDPSVFEV 720
 DB 661 YEKARLKEIMRIMGLDNLISLFWISLILPLVNSAGLLVILKIGNLLPYSDPSVFEV 720
 QY 721 FLVSFAVVTILQCFILISTFESRANLAAAGGIYFTLYLPYLCVAMODYVGTILKIFAS 780
 DB 721 FLVSFAVVTILQCFILISTFESRANLAAAGGIYFTLYLPYLCVAMODYVGTILKIFAS 780
 QY 781 LLSVPAFGCEYFALFEEOGIGVOMDLFESPYEEDGFMULTISISMILDFTLGYMTW 840
 DB 781 LLSVPAFGCEYFALFEEOGIGVOMDLFESPYEEDGFMULTISISMILDFTLGYMTW 840
 QY 841 YLEAFPGOYGIIPRWYPCYKSYWFGESDEKSHGPSNOKRMEICOMEDEPHLKIGVS 900
 DB 841 YLEAFPGOYGIIPRWYPCYKSYWFGESDEKSHGPSNOKRMEICOMEDEPHLKIGVS 900
 QY 901 IONLVKVRDGNKAVAVDGLALNFYEGQITTSFLGNGAGKTTNSILGLPPTSGAYIL 960
 DB 901 IONLVKVRDGNKAVAVDGLALNFYEGQITTSFLGNGAGKTTNSILGLPPTSGAYIL 960
 QY 961 GKDIREKSTIRONLGVCPQHNVLFDMLTVEHIMFYARLKGLSKHYAKAMEQMALDVG 1020
 DB 961 GKDIREKSTIRONLGVCPQHNVLFDMLTVEHIMFYARLKGLSKHYAKAMEQMALDVG 1020

QY 1021 LBSSKLKSTQSLSGGMOKRISVALAFVGSXVVLDEPTAGVDPYSRNGIWEILLKYRQ 1080
 DB 1021 LBSSKLKSTQSLSGGMOKRISVALAFVGSXVVLDEPTAGVDPYSRNGIWEILLKYRQ 1080
 QY 1081 GRTIILSTHHMEADVDLGRILAIISHKILCCVSSSLFKNOLGTGYTLTLVKKVESLS 1140
 DB 1081 GRTIILSTHHMEADVDLGRILAIISHKILCCVSSSLFKNOLGTGYTLTLVKKVESLS 1140
 QY 1141 SCRNSSVSYLKKEDSVSOSSDAGLSDSHESDPLTIDVSAISMLIKRHVSEARLVEDI 1200
 DB 1141 SCRNSSVSYLKKEDSVSOSSDAGLSDSHESDPLTIDVSAISMLIKRHVSEARLVEDI 1200
 QY 1201 GHELTLYLPEAKKEGAFVLFHEIDRLSDLGISISYGISETTLEIFLKAEEGVAE 1260
 DB 1201 GHELTLYLPEAKKEGAFVLFHEIDRLSDLGISISYGISETTLEIFLKAEEGVAE 1260
 QY 1261 TSDGTLPARNRRAFGKOSCLRPFTEDDAADPNDSIDIPESRETDLSGMCKGSYQVK 1320
 DB 1261 TSDGTLPARNRRAFGKOSCLRPFTEDDAADPNDSIDIPESRETDLSGMCKGSYQVK 1320
 QY 1321 GMLTLOOQFVALLKRLILARRSKGFFAOIYLPAYFVCIALVESLIVPFGKYPSLEQ 1380
 DB 1321 GMLTLOOQFVALLKRLILARRSKGFFAOIYLPAYFVCIALVESLIVPFGKYPSLEQ 1380
 QY 1381 PMYNEOYTFVSNDAPEDTGTLELNLATKDPGFGTCMGHNPIDTPOAGEEEMTAP 1440
 DB 1381 PMYNEOYTFVSNDAPEDTGTLELNLATKDPGFGTCMGHNPIDTPOAGEEEMTAP 1440
 QY 1441 VQOTIMDLFONGNMTNOMPSPACOCSSDKIKKMLPYCPGAGLPPORRONTADILQDL 1500
 DB 1441 VQOTIMDLFONGNMTNOMPSPACOCSSDKIKKMLPYCPGAGLPPORRONTADILQDL 1500
 QY 1501 TGRNLSIDYLVKTYVOIILAKSLKNKIWNEFRYGFSLGVSNTQALPPSOVYNAIKOMK 1560
 DB 1501 TGRNLSIDYLVKTYVOIILAKSLKNKIWNEFRYGFSLGVSNTQALPPSOVYNAIKOMK 1560
 QY 1561 HIKLAODSSADRFNLISLGFMTGLDTRNNVKWFENKGMHAISSEFLVJNNAILRANLQK 1620
 DB 1561 HIKLAODSSADRFNLISLGFMTGLDTRNNVKWFENKGMHAISSEFLVJNNAILRANLQK 1620
 QY 1621 GENPSHYGTAFENHPLNLTKQOLSEVALMTTSVDVLYSTCYIFAMSFPASFEVFLQER 1680
 DB 1621 GENPSHYGTAFENHPLNLTKQOLSEVALMTTSVDVLYSTCYIFAMSFPASFEVFLQER 1680
 QY 1681 VSKAKHLQFISGVKPYIYMLSNFVDMCNVVPATLVIILFICFOOKSVYSTNLPLVAL 1740
 DB 1681 VSKAKHLQFISGVKPYIYMLSNFVDMCNVVPATLVIILFICFOOKSVYSTNLPLVAL 1740
 QY 1741 LLLLYGMSITPLMYPASFVKIIPSTAYVVLTSVNLFIGINGSVATFVLELFTDNKLNNIN 1800
 DB 1741 LLLLYGMSITPLMYPASFVKIIPSTAYVVLTSVNLFIGINGSVATFVLELFTDNKLNNIN 1800
 QY 1801 DILKSVLIPPHFCGLGLDMYKNOAMADALERPENFVSPLSMDLVRGMLFMAAVER 1860
 DB 1801 DILKSVLIPPHFCGLGLDMYKNOAMADALERPENFVSPLSMDLVRGMLFMAAVER 1860
 QY 1861 VVFFLITVLYOYRFFIIPRPVNAKLSPINDEDEDVREBORIILDGGOINDILEIKELTKI 1920
 DB 1861 VVFFLITVLYOYRFFIIPRPVNAKLSPINDEDEDVREBORIILDGGOINDILEIKELTKI 1920
 QY 1921 YRRKRPAVDRIYVIGIPPEGCGLLGVNAGKSTFEMILGDTTVYRGDAFLKNSILSN 1980
 DB 1921 YRRKRPAVDRIYVIGIPPEGCGLLGVNAGKSTFEMILGDTTVYRGDAFLKNSILSN 1980
 QY 1981 THEVHONNGYOPOPATIELLTGREHVEFFALLRGVPEKEVKGWEMAIRKGLVYKGEK 2040
 DB 1981 THEVHONNGYOPOPATIELLTGREHVEFFALLRGVPEKEVKGWEMAIRKGLVYKGEK 2040
 QY 2041 YAGNTSGGNKRLSTAMALLGGPPVYVLEDEPTTGMDCARARFIMNCLASVYKREGSVLT 2100
 DB 2041 YAGNTSGGNKRLSTAMALLGGPPVYVLEDEPTTGMDCARARFIMNCLASVYKREGSVLT 2100
 QY 2101 SHSMECEALCTRMALIMVNGRFRCLGSVQHLKNRPGDYTTIVBRIAGSNPDLKPVODFFG 2160

Db 2101 SHSMEECALCTMA.LMVGRRCLSGVOH.KNREGDGYTVIRIAGSNPLK.PVODEFG 2160
 QY 2161 LAFPGSVLKEKRRNMLQYOLPSSLSLARIFSLSSQSKRIHIEYVSQTTLDQVYNF 2220
 Db 2161 LAFPGSVPEKRRNMLQYOLPSSLSLARIFSLSSQSKRIHIEYVSQTTLDQVYNF 2220
 QY 2221 AKQSDDDLKDLSTLHKNOTVDVAVLSFLQDEKVEKESYV 2261
 Db 2221 AKQSDDDLKDLSTLHKNOTVDVAVLSFLQDEKVEKESYV 2261

RESULT 2

ACBL_MOUSE

ID ACBL_MOUSE STANDARD: PRT: 2261 AA.

AC P41233:

DT 01-FEB-1995 (Rel. 31, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE ATP-binding cassette, sub-family A, member 1 (ATP-binding cassette transporter 1) (ATP-binding cassette 1) (ABC-1).

GN ABCA1 OR ABCA1.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=DNA/2; TISSUE=Macrophage;

RA MEDLINE=94375008; PubMed=8088782;

RT Luciani M.F., Denizot F., Savary S., Mattei M.-G., Chimini G.;

RT "Cloning of two novel ABC transporters mapping on human chromosome 9."

RT Genomics 21:150-159(1994).

RL [2]

RN SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J;

RA Qiu Y., Cavelier L., Chiu S., Rubin E., Cheng J.-F.;

RT "Human and mouse ABCA1 comparative sequencing and transgenesis studies identify potential regulatory sequences."

RT Submitted (Jul-2000) to the EMBL/Genbank/DBJ databases.

RL -1- FUNCTION: CAMP-DEPENDENT AND SULFONYLUREA-SENSITIVE ANION TRANSPORTER. KEY GATEKEEPER INFLUENCING INTRACELLULAR CHOLESTEROL TRANSPORT (BY SIMILARITY).

CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED IN ADULT TISSUES. HIGHEST LEVELS ARE FOUND IN PREGNANT UTERUS AND UTERUS.

CC -1- DOMAIN: MULTIFUNCTIONAL POLYPEPTIDE WITH TWO HOMOLOGOUS HALVES, EACH CONTAINING AN HYDROPHOBIC MEMBRANE-ANCHORING DOMAIN AND AN -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCA SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial or send an email to license@isb-stb.ch).

 EMBL: X75926; CAAG3530.1; ALT_INIT.
 EMBL: AF287263; AAG39073.1; ALT_INIT.
 DR M5D; MG1:99607; Abca1.
 DR InterPro: IPR003593; AAA_ATPase.
 DR InterPro: IPR003439; ABC_transporter.
 DR Pfam: PF00005; ABC_tran; 2.
 DR Prodom: PD000006; ABC_transporter; 2.
 DR SMART: SM00382; AAA; 1.
 DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
 KM ATP-binding; Glycoprotein; Transmembrane; Transport.
 FT TRANSMEM 26 42
 FT TRANSMEM 640 656 POTENTIAL.
 FT TRANSMEM 690 706 POTENTIAL.
 FT TRANSMEM 717 733 POTENTIAL.

FT TRANSMEM 749 765 POTENTIAL.
 FT TRANSMEM 771 787 POTENTIAL.
 FT TRANSMEM 1041 1057 POTENTIAL.
 FT TRANSMEM 1351 1367 POTENTIAL.
 FT TRANSMEM 1661 1677 POTENTIAL.
 FT TRANSMEM 1708 1724 POTENTIAL.
 FT TRANSMEM 1737 1753 POTENTIAL.
 FT TRANSMEM 1775 1791 POTENTIAL.
 FT TRANSMEM 1854 1870 POTENTIAL.
 FT NP_BIND 933 940
 FT NP_BIND 1946 1953
 FT CARBOHYD 14 14
 FT CARBOHYD 98 98
 FT CARBOHYD 151 151
 FT CARBOHYD 161 161
 FT CARBOHYD 196 196
 FT CARBOHYD 244 244
 FT CARBOHYD 292 292
 FT CARBOHYD 337 337
 FT CARBOHYD 349 349
 FT CARBOHYD 400 400
 FT CARBOHYD 478 478
 FT CARBOHYD 489 489
 FT CARBOHYD 521 521
 FT CARBOHYD 820 820
 FT CARBOHYD 1144 1144
 FT CARBOHYD 1294 1294
 FT CARBOHYD 1453 1453
 FT CARBOHYD 1499 1499
 FT CARBOHYD 1504 1504
 FT CARBOHYD 1637 1637
 FT CARBOHYD 2044 2044
 FT CARBOHYD 2238 2238
 FT CONFLICT 1567 1568
 FT CONFLICT 2024 2024
 SQ SEQUENCE 2261 AA; 254011 MW; FA62B21FD1D09F9 CRC64;

Query Match 95.4%; Score 11256; DB 1; Length 2261;
 Best Local Similarity 95.0%; Pred. No. 0;
 Matches 2149; Conservative 55; Mismatches 57; Indels 0; Gaps 0;

QY 1 MACMQLRLMLKNTFRRTQCLLEVPMPFLILISVRLSPYEDHCHFPNKA 60
 Db 1 MACMQLRLMLKNTFRRTQCLLEVPMPFLILISVRLSPYEDHCHFPNKA 60
 QY 61 MPSAGTLPWVGIIICANNPCFRYPFGAPGVGNKSTIVARLSDARLLYSQKT 120
 Db 61 MPSAGTLPWVGIIICANNPCFRYPFGAPGVGNKSTIVARLSDARLLYSQKT 120
 QY 121 SMKMRKVLFTLQOIKSSNKLQDPLVNEFSEFLYHNLSPKSTVDKMLRADVLH 180
 Db 121 SIKMHVLRMLRKIKRPNLNKLQDPLVNEFSEFLYHNLSPKSTVDKMLRADVLH 180
 QY 181 KVFLOGVLHLTSCNSKSEMIQDQEVSEFLGCPREKILAAERVLRSMQDLKYL 240
 Db 181 KVFLOGVLHLTSCNSKSEMIQDQEVSEFLGCPREKILAAERVLRSMQDLKYL 240
 QY 241 RLINSTSPSPSKELAEFTYTLHSLGTLAQELFSRMSQDMRQEVMPFLJNVSSSSSTOI 300
 Db 241 RLINSTSPSPSKELAEFTYTLHSLGTLAQELFSRMSQDMRQEVMPFLJNVSSSSSTOI 300
 QY 301 YQAVSRIVCGHPGGLKIKSLNMYEDNNYKALFGCNGCEDEAFEPYNSSTPPCNDLMK 360
 Db 301 YQAVSRIVCGHPGGLKIKSLNMYEDNNYKALFGCNGCEDEAFEPYNSSTPPCNDLMK 360
 QY 361 NLESSPLSLIKKALKPLLVGLILTPPTPATROYMAEVNKTFOELAVFHDLEGMSELS 420
 Db 361 NLESSPLSLIKKALKPLLVGLILTPPTPATROYMAEVNKTFOELAVFHDLEGMSELS 420
 QY 421 PKIWTFMNSQMDLVRLMLDSRDNDHFWEQDGLDMWTADDIYAFILAKHPEDVOSSNGS 480
 Db 421 PKIWTFMNSQMDLVRLMLDSRDNDHFWEQDGLDMWTADDIYAFILAKHPEDVOSSNGS 480

QY 481 VYTWREAFNETNOAIRTISREMECVNLKLEPLATEVWLKNSMELDERKFMAGIVFTG 540
 |||||
 Db 481 VYTWREAFNETNOAIQTTSRMECVNLKLEPLATEVWLKNSMELDERKFMAGIVFTG 540
 |||||
 QY 541 IFPGSIELPHHVKYKIRMDIDWERTNKIKDGYWDPGPRADEFEDMKRYWGGFAYLDVY 600
 |||||
 Db 541 ITDSEVELPHHVKYKIRMDIDWERTNKIKDGYWDPGPRADEFEDMKRYWGGFAYLDVY 600
 |||||
 QY 601 EQAIIIVLTGTEKKTGYVMOQPRPCYVDITFLVMSRSMPLFMTLAMYSAVATKIGIV 660
 |||||
 Db 601 EQAIIIVLTGTEKKTGYVMOQPRPCYVDITFLVMSRSMPLFMTLAMYSAVATKIGIV 660
 |||||
 QY 661 YEKERARKETMRIMGDNGITLMSFWSSILPLVSAGLIVILKGNLLPYSDPSVYFV 720
 |||||
 Db 661 YEKERARKETMRIMGDNGITLMSFWSSILPLVSAGLIVILKGNLLPYSDPSVYFV 720
 |||||
 QY 721 FLVFAVAVTILLOCFILISTFLSRANLAAAGGIIYFTLYLPYLVCAVMOQYVFTLKIFAS 780
 |||||
 Db 721 FLVFAVAVTILLOCFILISTFLSRANLAAAGGIIYFTLYLPYLVCAVMOQYVFTLKIFAS 780
 |||||
 QY 781 LLSVAFGFGCEYFALFEEOGIGVOMDNLFESEVEEDGFNLTTISIMLEFDTFLYGVMTW 840
 |||||
 Db 781 LLSVAFGFGCEYFALFEEOGIGVOMDNLFESEVEEDGFNLTTISIMLEFDTFLYGVMTW 840
 |||||
 QY 841 YIEAVFPQGYGIPRPWYFPCTKSYWGESEDEKSHPGSNOKRMESEICMEEPHILKIGVS 900
 |||||
 Db 841 YIEAVFPQGYGIPRPWYFPCTKSYWGESEDEKSHPGSNOKRMESEICMEEPHILKIGVS 900
 |||||
 QY 901 IONLVKVRDGMKVAVDGLANFEGGITSFLGNAGAKTTTMSIITGLPPTSGTAYIL 960
 |||||
 Db 901 IONLVKVRDGMKVAVDGLANFEGGITSFLGNAGAKTTTMSIITGLPPTSGTAYIL 960
 |||||
 QY 961 GKDISEMSTIRONLGVCOHNVLFDMTYVEHIMFYARLKIGSEKRYVAKEMQALDVG 1020
 |||||
 Db 961 GKDISEMSTIRONLGVCOHNVLFDMTYVEHIMFYARLKIGSEKRYVAKEMQALDVG 1020
 |||||
 QY 1021 LPSSKLSKTSOLSGOMOKLSVALAFYGGSKVILDEPTAGVPRYRGRIGWELLKYRQ 1080
 |||||
 Db 1021 LPSSKLSKTSOLSGOMOKLSVALAFYGGSKVILDEPTAGVPRYRGRIGWELLKYRQ 1080
 |||||
 QY 1081 GRTIILSTHMHMEDADVLGRILAIISHGRLCCVSSSLFKNLGTYLYTLVKKVESLS 1140
 |||||
 Db 1081 GRTIILSTHMHMEDADVLGRILAIISHGRLCCVSSSLFKNLGTYLYTLVKKVESLS 1140
 |||||
 QY 1141 SCRNSSTVSYLKEDSVSSSDAGLGDHSDPDLTIDVSAISLIRKHYSEARLYEDI 1200
 |||||
 Db 1141 SCRNSSTVSYLKEDSVSSSDAGLGDHSDPDLTIDVSAISLIRKHYSEARLYEDI 1200
 |||||
 QY 1201 GHLEITVLPPEAKKEGAFVELFHEIDRLSDIGISSYGISETTLEIFLKAEEGVDAE 1260
 |||||
 Db 1201 GHLEITVLPPEAKKEGAFVELFHEIDRLSDIGISSYGISETTLEIFLKAEEGVDAE 1260
 |||||
 QY 1261 TSDGTLPARNRRAFGDKOSCLPTEEDADPNDSIDPESRENDILSGMKSGSYOVK 1320
 |||||
 Db 1261 TSDGTLPARNRRAFGDKOSCLPTEEDADPNDSIDPESRENDILSGMKSGSYOVK 1320
 |||||
 QY 1321 GMRKLTQOQFVAILMKRLLIRSRKGFPAQIYPAVFCIALVFSLIYPPFGKPSLEIQ 1380
 |||||
 Db 1321 GMRKLTQOQFVAILMKRLLIRSRKGFPAQIYPAVFCIALVFSLIYPPFGKPSLEIQ 1380
 |||||
 QY 1381 PMWYNEQYTVSNDAPEDTGTLELALNATKDPGFGTRCHMGKNIPPTPOCAEEFMTAP 1440
 |||||
 Db 1381 PMWYNEQYTVSNDAPEDTGTLELALNATKDPGFGTRCHMGKNIPPTPOCAEEFMTAP 1440
 |||||
 QY 1441 VPQITMDLFONGNMTQNPSPACOCSSDKTKKMLPVCPPGAGLIPPRORONTADILLOD 1500
 |||||
 Db 1441 VPQITMDLFONGNMTQNPSPACOCSSDKTKKMLPVCPPGAGLIPPRORONTADILLOD 1500
 |||||
 QY 1501 TGRNISDLVKTYYOQIITAKSLKNKIWNEFRYGGFSLGVSNTQALPPSOEVNATIKOMK 1560
 |||||
 Db 1501 TGRNISDLVKTYYOQIITAKSLKNKIWNEFRYGGFSLGVSNTQALPPSOEVNATIKOMK 1560
 |||||
 QY 1561 HLKLAKSSADRFNLNSLGRFMTGLDTRNNVYVWFNNKGWHAISSEFLVNNAILRANLOK 1620
 |||||

Db 1561 LKLTQTSADRFNLNSLGRFMTGLDTRNNVYVWFNNKGWHAISSEFLVNNAILRANLOK 1620
 |||||
 QY 1621 GENPSHYGITAENHPIINLTQKOLSEVALMTTSVDVYVISCYIFAMSPVPASFFVLLQER 1680
 |||||
 Db 1621 GENPSHYGITAENHPIINLTQKOLSEVALMTTSVDVYVISCYIFAMSPVPASFFVLLQER 1680
 |||||
 QY 1681 VSKAKILOFTSGVCPVITYLSNFWDMCNVYVPAVLVITIFICFOOKSYSSITNLPVAL 1740
 |||||
 Db 1681 VSKAKILOFTSGVCPVITYLSNFWDMCNVYVPAVLVITIFICFOOKSYSSITNLPVAL 1740
 |||||
 QY 1741 LLLYGMSTIPLMPASFEVKIPSTAYVLTISVNLFIGINGSVATFVLELFTMDKNNIN 1800
 |||||
 Db 1741 LLLYGMSTIPLMPASFEVKIPSTAYVLTISVNLFIGINGSVATFVLELFTMDKNNIN 1800
 |||||
 QY 1801 DILKSVLIFPHFCIGRLIDWKYNOAMADALEREGENFVSPLSMDLVGRNLFMAAVEG 1860
 |||||
 Db 1801 DILKSVLIFPHFCIGRLIDWKYNOAMADALEREGENFVSPLSMDLVGRNLFMAAVEG 1860
 |||||
 QY 1861 VVFLITVLIQYRFFIRPREVNAKLSPUNDEDEDVREBQRIIDGGQNDILEIKELTKI 1920
 |||||
 Db 1861 VVFLITVLIQYRFFIRPREVNAKLSPUNDEDEDVREBQRIIDGGQNDILEIKELTKI 1920
 |||||
 QY 1921 YRRKRPAVDRIYGVIPGECFGLGVNGAGKSTFKMLTGDVTYRGDAFLNKSILSN 1980
 |||||
 Db 1921 YRRKRPAVDRIYGVIPGECFGLGVNGAGKSTFKMLTGDVTYRGDAFLNKSILSN 1980
 |||||
 QY 1981 IHEHONNGYCPQDQDAITELTIGREHEVEFALLRGVEKEVKGEMAIRKGLYKYGK 2040
 |||||
 Db 1981 IHEHONNGYCPQDQDAITELTIGREHEVEFALLRGVEKEVKGEMAIRKGLYKYGK 2040
 |||||
 QY 2041 YAGNYSNGNRKRLSTAMALJIGPPVYFLDEPTGMDKARFLMNCALSYVKEGRSVLT 2100
 |||||
 Db 2041 YAGNYSNGNRKRLSTAMALJIGPPVYFLDEPTGMDKARFLMNCALSYVKEGRSVLT 2100
 |||||
 QY 2101 SHSMECEALCTRNAIMVNGRFRCLGSVOHLKNGFGDGYTVRIAGSNPDLKPVODFG 2160
 |||||
 Db 2101 SHSMECEALCTRNAIMVNGRFRCLGSVOHLKNGFGDGYTVRIAGSNPDLKPVODFG 2160
 |||||
 QY 2161 LAEPGSVLEKERRNMLQYQLPSSJSLARITSLISQSKRLHIEDYSVQTTLDQVFN 2220
 |||||
 Db 2161 LAEPGSVLEKERRNMLQYQLPSSJSLARITSLISQSKRLHIEDYSVQTTLDQVFN 2220
 |||||
 QY 2221 AKDQSDHKLKDLSTLHKNOTVAVVAVLFSPLQDEKVESYV 2261
 |||||
 Db 2221 AKDQSDHKLKDLSTLHKNOTVAVVAVLFSPLQDEKVESYV 2261
 |||||

RESULT 3

ID ABR_C_HUMAN STANDARD: PRT: 2273 AA.
 AC P78363; 060438; 060915; 015112;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Retinal-specific ATP-binding cassette transporter (RIM ABC
 DE transporter) (RIM protein) (RMP) (Stargardt disease protein).
 GN ABCA4 OR ABR.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OC NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A., VARIANTS STGD, AND VARIANTS HIS-846 AND GLN-943.
 RX MEDLINE-97207641; PubMed=9054934;
 RA Allikmets R., Singh N., Sun H., Shroyer N.F., Hutchinson A.,
 RA Childambaram A., Gettrard B., Baird L., Stauffer D., Peiffer A.,
 RA Ratner A., Smallwood P.M., Li Y., Anderson K.L., Lewis R.A.,
 RA Nathans J., Leppert M., Dean M., Jabski J.R.,
 RA "A photoreceptor cell-specific ATP-binding transporter gene (ABCR) is
 RT mutated in recessive Stargardt macular dystrophy."
 RL Nat. Genet. 15:236-246(1997).
 RN [2]

RP SEQUENCE FROM N.A.
RX MEDLINE=97345663; PubMed-9202155;
RA Azarian S.M., Travis G.H.;
RT "The photoreceptor rim protein is an ABC transporter encoded by the
RL gene for recessive Stargardt's disease (ABCR).";
RN FEBS Lett. 409:247-252(1997).
[3]
RP SEQUENCE FROM N.A., AND VARIANTS STGD TRP-18 AND CYS-212.
RX MEDLINE=98163759; PubMed-9503029;
RA Gerber S., Rozet J.-M., van de Pol T.J.R., Hoyng C.B., Munnich A.,
RA Blankenagel A., Kaplan J., Cremers F.P.M.;
RT "Complete exon-intron structure of the retina-specific ATP binding
RT transporter gene (ABCR) allows the identification of novel mutations
RL underlying Stargardt disease.";
RN Genomics 48:139-142(1998).
[4]
RP SEQUENCE FROM N.A., AND VARIANTS STGD.
RX MEDLINE=98141123; PubMed-9490294;
RA Nasonkin I., Tilling M., Koehler M.R., Schmidt M., Molday R.S.,
RA Weber B.H.F.;
RT "Mapping of the rod photoreceptor ABC transporter (ABCR) to 1p21-p22.1
RT and identification of novel mutations in Stargardt's disease.";
RN Hum. Genet. 102:21-26(1998).
[5]
RP CHARACTERIZATION.
RX MEDLINE=99175213; PubMed=10075733;
RA Sun H., Molday R.S., Nathans J.;
RT "Retinal stimulates ATP hydrolysis by purified and reconstituted ABCR,
RT the photoreceptor-specific ATP-binding cassette transporter
RL responsible for Stargardt disease.";
RN J. Biol. Chem. 274:8269-8281(1999).
[6]
RP DISEASE.
RX MEDLINE=98133912; PubMed=9466990;
RA Cremers F.P.M., van de Pol D.J.R., van Driel M.A., den Hollander A.I.,
RA van Haren F.J.J., Koeber N.V.A.M., Tijmes N., Bergen A.A.B.,
RA Rohrschneider K., Blankenagel A., Plinckers A.O.L.G., Deutman A.F.,
RA Hoyng C.B.;
RT "Autosomal recessive retinitis pigmentosa and cone-rod dystrophy
RT caused by splice site mutations in the Stargardt's disease gene
RL ABCR.";
RN Hum. Mol. Genet. 7:355-362(1998).
[7]
RP VARIANTS ARM2, AND VARIANTS.
RX MEDLINE=97442530; PubMed=9295268;
RA Allikmets R., Shroyer N.F., Singh N., Seddon J.M., Lewis R.A.,
RA Bernstein P.S., Pfeiffer A., Zabitskij N.A., Li Y., Hutchinson A.,
RA Dean M., Lupski J.R., Leppert M.;
RT "Mutation of the Stargardt disease gene (ABCR) in age-related macular
RT degeneration.";
RN Science 277:1805-1807(1997).
[8]
RP VARIANTS STGD W-18; C-212; H-636; M-1019; V-1038; C-1108; W-1640;
RP S-1977 AND H-2107, AND VARIANTS FEM P-11; P-541; V-1038; F-1091;
RP C-1508; F-1970 AND R-1971.
RX MEDLINE=98454319; PubMed=9781034;
RA Rozet J.-M., Gerber S., Souled E., Perrault I., Chatelain S., Ghazi I.,
RA Leowski C., Dullier J.-L., Munnich A., Kaplan J.;
RT "Spectrum of ABCR gene mutations in autosomal recessive macular
RT dystrophies.";
RN Eur. J. Hum. Genet. 6:291-295(1998).
[9]
RP VARIANTS STGD.
RX MEDLINE=99138655; PubMed=9973280;
RA Lewis R.A., Shroyer N.F., Singh N., Allikmets R., Hutchinson A.,
RA Li Y., Lupski J.R., Leppert M., Dean M.;
RT "Genotype/phenotype analysis of a photoreceptor-specific ATP-binding
RT cassette transporter gene, ABCR, in Stargardt disease.";
RN Am. J. Hum. Genet. 64:422-434(1999).
[10]
RP VARIANTS STGD, AND VARIANTS.
RX MEDLINE=99192348; PubMed=10090887;
RA Maugeri A., van Driel M.A., van de Pol D.J.R., Klievering B.J.,
RA van Haren F.J.J., Tijmes N., Bergen A.A.B., Rohrschneider K.,
RA Blankenagel A., Plinckers A.J.L.G., Dahl N., Brunner H.G.,
RA Deutman A.F., Hoyng C.B., Cremers F.P.M.;
RT "The 2588G->C mutation in the ABCR gene is a mild frequent founder
RT mutation in the western European population and allows the
RT classification of ABCR mutations in patients with Stargardt disease.";
RN Am. J. Hum. Genet. 64:1024-1035(1999).
[11]
RP VARIANT STGD TYR-54, AND VARIANT ALA-863.
RX MEDLINE=20077755; PubMed=10612508;
RA Zhang K., Garibaldi D.C., Kniazeva M., Albini T., Chiang M.F.,
RA Kerrigan M., Sunness J.S., Han M., Allikmets R.;
RT "A novel mutation in the ABCR gene in four patients with autosomal
RT recessive Stargardt disease.";
RN Am. J. Ophthalmol. 128:720-724(1999).
[12]
RP VARIANTS STGD V-60; R-206; N-300; P-541; A-849; P-974; V-1038; C-1108;
RP L-1408; R-1488; D-1652; P-1729; E-1961; W-2038; W-2077; H-2107; R-2128
RP AND Y-2150.
RX MEDLINE=99221420; PubMed=10206579;
RA Fishman G.A., Stone E.M., Grover S., Derlacki D.J., Haines H.L.,
RA Hockey R.R.;
RT "Variation of clinical expression in patients with Stargardt dystrophy
RT and sequence variations in the ABCR gene.";
RN Arch. Ophthalmol. 117:504-510(1999).
[13]
RP VARIANTS GLU-1961 AND ASN-2177.
RX MEDLINE=20349288; PubMed=10880298;
RA Allikmets R., Tammur J., Hutchinson A., Lewis R.A., Shroyer N.F.,
RA Dalakishvili K., Lupski J.R., Steiner K., Paulikhoff D., Holz F.G.,
RA Singh N., Pfeiffer A., Zabitskij N.A., Leppert M., Seddon J.M.,
RA Zhang K., Sunness J.S., Udar N.S., Yelchits S., Silva-Garcia R.,
RA Small K.W., Simonelli F., Testa F., D'Urso M., Brancato R.,
RA Rinaldi E., Iyegust S., Taube A., Wadelius C., Souled E., Ducrocq D.,
RA Kaplan J., Assink J.J.M., ten Brink J.B., de Jong P.T.V.M.,
RA Bergen A.A.B., Maugeri A., van Driel M.A., Hoyng C.B., Cremers F.P.M.,
RA Palma E., Coco R., Balcells S., Gonzalez-Duarte R., Kermani S.,
RA Stanga P., Bhattacharya S.S., Bird A.C.;
RT "Further evidence for an association of ABCR alleles with age-related
RT macular degeneration.";
RN Am. J. Hum. Genet. 67:487-491(2000).
[14]
RP VARIANTS STGD E-60; T-60; E-65; L-68; R-72; C-212; S-230; S-247;
RP V-328; K-471; P-541; Q-572; R-607; K-635; C-653; Y-764; R-765; A-901;
RP I-959; K-1036; V-1038; P-1063; D-1087; C-1097; C-1108; L-1380; K-1399;
RP P-1430; V-1440; H-1443; L-1486; Y-1488; M-1537; P-1689; L-1705;
RP T-1733; R-1748; P-1763; R-1885; H-1898; A-1961; S-1977; G-2077
RP W-2077 AND V-2241, AND VARIANTS Q-152; H-212; R-423; I-552; R-914;
RP Q-943; T-1562; I-1868; M-1921; L-1948; F-1970; A-2059; N-2177 AND
RP V-2216.
RX MEDLINE=20442027; PubMed=10958763;
RA Rivera A., White K., Stoehr H., Steiner K., Hemmrich N., Grimm T.,
RA Jurkles B., Lorenz B., Scholl H.P.N., Apfelstedt-Sylla E.,
RA Weber B.H.F.;
RT "A comprehensive survey of sequence variation in the ABC4 (ABCR) gene
RT in Stargardt disease and age-related macular degeneration.";
RN Am. J. Hum. Genet. 67:800-813(2000).
[15]
RP VARIANTS CORD3 GLU-65; CYS-212; PRO-541; ALA-863; GLY-863 DEL;
RP VAL-1038; LYS-1122; TYR-1490 AND ASP-1598.
RX MEDLINE=20442040; PubMed=10958761;
RA Maugeri A., Klievering B.J., Rohrschneider K., Blankenagel A.,
RA Brunner H.G., Deutman A.F., Hoyng C.B., Cremers F.P.M.;
RT "Mutations in the ABC4 (ABCR) gene are the major cause of autosomal
RT recessive cone-rod dystrophy.";
RN Am. J. Hum. Genet. 67:960-966(2000).
[16]
RP VARIANTS STGD ASP-340; GLN-572; ALA-863; SER-965; VAL-1038; ALA-1780
RP AND HIS-1898, AND VARIANT GLN-943.
RX MEDLINE=20208356; PubMed=10746567;
RA Shroyer N.F., Lewis R.A., Lupski J.R.;
RT "Complex inheritance of ABCR mutations in Stargardt disease: linkage

RT disequilibrium, complex alleles, and pseudodominance";
 RL Hum. Genet. 106:244-248(2000).
 RN [17]
 RX VARIANTS STGD
 MEDLINE=20098082; PubMed=10634594;
 RA Papaioannou M., Ocaña L., Bessant D., Lois N., Bird A.C., Payne A.,
 RA Bhatnagar S.S.;
 RT "An analysis of ABCR mutations in British patients with recessive
 RT retinal dystrophies";
 RL Invest. Ophthalmol. Vis. Sci. 41:16-19(2000).
 RN [18]
 RX VARIANTS STGD C-212; D-767; T-897; V-1038; K-1087; K-1399; Q-1640 AND
 R E-1961, AND VARIANT HIS-212.
 RX MEDLINE=20174852; PubMed=10711710;
 RA Simionelli F., Testa F., de Crechto G., Rinaldi E., Hutchinson A.,
 RA Atkinson A., Dean M., D'urso M., Allikmets R.;
 RT "New ABCR mutations and clinical phenotype in Italian patients with
 RT Stargardt disease";
 RL Invest. Ophthalmol. Vis. Sci. 41:892-897(2000).
 RN [19]
 RX CHARACTERIZATION OF VARIANTS, AND MUTAGENESIS OF GLY-966; IYS-969;
 RP GLY-1975 AND IYS-1978.
 RX MEDLINE=20472331; PubMed=11017087;
 RA Sun H., Smallwood P.M., Nathans J.;
 RT "Biochemical defects in ABCR protein variants associated with human
 RT retinopathies";
 RL Nat. Genet. 26:242-246(2000).
 RN [20]
 RX VARIANT STGD ASN-972, AND VARIANTS GLN-943; ILE-1868 AND LEU-1948.
 RX MEDLINE=21478761; PubMed=1159493;
 RA Eksandh L., Ekstrom U., Abrahamson M., Bauer B., Andreasson S.;
 Query Match 49.8%; Score 5870.5; DB 1; Length 2273;
 Best Local Similarity 50.0%; Pred. No. 0;
 Matches 1165; Conservative 358; Mismatches 635; Indels 171; Gaps 27;

DB 522 LDKFESYNDENQULQORALSLILENNMFAGVFPDMYPMWTSLLPBYKXKIRMDIDVYERT 581
 567 NKIKQYWDPCGRADPEEDMDKRYWNGFAYLQDVQOALIVVLGTEKKRGVYQWQMPYC 626
 582 NKIKRYWDSGFRADPEVDFRIRWGFAYLQDVQOQITRSQVQAEAPVQIYIQOAPYPC 641
 627 YVDDIFLRYMSRMPLEFMTLAMIYSAVITIKGIYVEAKRLKEMRIMGLDNLMSWF 666
 642 EYDSEFMILNCFIFPVLAMIVSVSMYKSYLEKRLKELTKNOGVSNAVIMCTWF 701
 687 ISLPLVLSAGLIVLILKGLNLPYSDPVEVFLSVFAVVTLLQCLISTLFSRANLA 746
 702 LQSFISMSISIFLTFIMHGRILHYSDPFLFLFLAFSTATIMLCFLSTFEFSKASLA 761
 747 AACGGIITYTLPLVYLCVAMQDYGFLTKFASLSLVAFCGCEAFALFEEOGIVQW 806
 762 AACGVITYFLPLPHLCFAWQDMRTAEKRAVSLSLSVAFGFGTEYLYVREBQGLQW 821
 807 DNLFSPEVERDGFNLTSTSMILPFTLYGVMTWYIEAVFGQYGPVYFPCTKSYF 866
 822 SNIGNSFTGEDERSFLSMQMLLAAYVGLAWYLDQVFPDQYGPVYFLLQESYWL 881
 867 G-----ESDEKSHPGSNOKRMSFICMEEFTHLKLGVSIONLVKY 907
 882 GCGCSTRERALEKTEPLETEDEPHD---EGIHDSFEREHPGVPGVCYANLVKI 937
 908 YRQGMKAVVNDGLNLFYEGQITSFILHNGCAGKTMTSITLGTLPSTAYILKQIRSE 967
 938 FEPGCRPAVRLNITTYENDITFELHNGAGKTTLSITLGLPPTSGVLYAGRDIFTS 997
 968 MSTIRQNLGVCPQHNVLFDMLVVEEHIWYARILKISEKHVYAEBOALDVGLPSSKLK 1027
 998 LDAVRQSLGMCPOHNLIFHLITVAEHMLFYAQIKGKSQEOALEMEAMLEDGL-HHNRN 1056
 1028 SKTSQLSGQNRKLSVALAVYGGSKVYIIDEPAGVDPYSRSGIEMELLYKRGQTIILS 1087
 1057 EEAQDLGGQNRKLSVALAVYGAQVYIIDEPISGVDPYSRSGIEMELLYKRGQTIILS 1116
 1088 THHNDADVLADRIATLISHGKLCVGSFLKNOQCTGYVLLYKRVESLSISCRNSS 1147
 1117 THHNDADVLADRIATLISHGKLCVGSFLKNOQCTGYVLLYKRVESLSISCRNSS 1175
 1148 TVSYLKEDSVSSQSDAGLCSHESDTLTIDVSALSNLRKVSARARVDEIGHELTV 1207
 1176 TCSGSSKQFS---TTCPAHVDLTPPEOVLDGVNEMDVVLAHVPKALVEGICQELIFL 1232
 1208 LPEFAKGCAGVLEFHEIDRLSDLSISYGISSETTLEELFLKVAEESGVDAETSDGLP 1267
 1233 LPNKFKAIRAVASLFRLEETLADLGISFQISDTPLEELFLKVTEDSDGFLFAGG--- 1289
 1268 ARNRRAFQDQSCLEPPTEDDAADPNDSDI-----DPESRTDLISGM 1311
 1290 AOKKRENNVPRHPCIGP-REKAGQTPODSNVCSFGAPAHAREGQPPPEPCPGQINTGT 1348
 1312 DKGSGYQKQKMLIQOQVVALMLKRLILARRSRKGFQVYLPAVFAVCIALVSLIVPFP 1371
 1349 -----OLVLIQVALLVLRFOHTIRSHKDFLAQVILPRTVFLALMISYIIPPE 1397
 1372 GYPSLELQPMWYNEQYTFVSNDAPEDTGTELLNALTKDPGEGTCRMESNP1PDTPCOA 1431
 1398 GREPALTLHPWLYGQYEFESMDPEGSQFTVYLDVILNRPFGFNRLKGCMBREYPC-G 1456
 1432 GDEBWTAPVOTYMDLFQNGMTMONSPPACOCSSQKIKKMLPVPYCGAGGPPPPORQ 1491
 1457 NSTPMKTPSPVSNITQLQOKQWQVNPSPSCROSTRKLTMLDECBGAGGPPPPORQ 1516
 1492 NTAADITODTGNISDYLVTQVQIILAKSLNKNKIVNEFFRGGFSLGVSNTQALPSEOEV 1551
 1517 RSTELIQLDLDKRNIDELVKTYPALIRSKLSKFPVNNQRYGGISIG---GKLPVVPIT 1572
 1552 NDAIKQMKHKLAKDASDRFLNSLGR-----FMICLDTRNNVVKV 1592

Db 1573 GEAIV-----GFLSLGRIMNVSQSPITREASKETLPDLKLEEDNIKV 1617
 QY 1593 WFNKNGMHAISFLNINNAIRALQGENDSHGITAENHPINLAKOOLSEVALMTTS 1652
 Db 1618 WFNKNGMHAISFLNINNAIRALQGENDSHGITAENHPINLAKOOLSEVALMTTS 1677
 QY 1653 VDLVLSICVIFAMSFVSPAFVFLVQERVSRAKHOFISGKPKVITVLSNVMCMKYV 1712
 Db 1678 VAAVAVACVIFMSFVSPAFVFLVQERVSRAKHOFISGKPKVITVLSNVMCMKYV 1737
 QY 1713 PATVITVITFCQOKSYVSNLPLVALLLLYGMSTITPLMYPASFEVKIIPSTAVVLTSS 1772
 Db 1738 SAGLVWVGIFGOKKAYSPENLPALVALLLYGMSTITPLMYPASFEVKIIPSTAVVLTSS 1797
 QY 1773 VNLFGINGSVATFVLELEFDNK-LNNINDLKSVPILFPEFCGRLLDMVKQOMADA 1831
 Db 1798 ANLFGINGSVATFVLELEFDNK-LNNINDLKSVPILFPEFCGRLLDMVKQOMADA 1857
 QY 1832 IERFENRNVSPDLNDVGRNLFMAAVEGVFLITVLIQYRFRIRPVNAKLSPLNDE 1891
 Db 1858 YAREGEHSANPFFHDLIGKMLFMAVEGVFLITVLIQYRFRIRPVNAKLSPLNDE 1917
 QY 1892 DEDVREERORITDGGGNDILEIKELIKIRRRKRPVNDICVIGIPGECFGLGVNGAG 1951
 Db 1918 DDDVAEEFORITDGGGNDILEIKELIKIRRRKRPVNDICVIGIPGECFGLGVNGAG 1977
 QY 1952 KSTFFKMLTGDPTVTRGDAPLAKNSILSNHEVQNGYCPQDADLDELITGHEHEFFA 2011
 Db 1978 KSTFFKMLTGDPTVTRGDAPLAKNSILSNHEVQNGYCPQDADLDELITGHEHEFFA 2037
 QY 2012 LLRGVPEKVGKGMALIKRLGLVKEKAYAGNSGNNRKLSTAMALIGPPVFLDEP 2071
 Db 2038 LLRGVPEKVGKGMALIKRLGLVKEKAYAGNSGNNRKLSTAMALIGPPVFLDEP 2097
 QY 2072 TTGMDFPARRFLMNCALSVYKGRSVLTSHSMECEALCTRALIVNGRFGCLSYOHL 2131
 Db 2098 TTGMDFPARRFLMNCALSVYKGRSVLTSHSMECEALCTRALIVNGRFGCLSYOHL 2157
 QY 2132 KNRGDCYITIVPLAGSN---PDLKPVQDFGLGAPGSLVLEKGRNMLQYOLPSLSL 2187
 Db 2158 KNRGDCYITIVPLAGSN---PDLKPVQDFGLGAPGSLVLEKGRNMLQYOLPSLSL 2215
 QY 2188 ARFESILSOSKRRHIEDSVSQTIDQYEVNPAKQSDHDLKLSLH 2236
 Db 2216 ARFESILSOSKRRHIEDSVSQTIDQYEVNPAKQSDHDLKLSLH 2261

RESULT 4
 ABC2_HUMAN
 ID ABC2_HUMAN STANDARD: PRT: 2436 AA.
 AC Q9BZC7;
 DT 16-OCT-2001 (rel. 40, Created)
 DT 16-OCT-2001 (rel. 40, Last sequence update)
 DT 15-JUN-2002 (rel. 41, Last annotation update)
 DE ATP-binding cassette, sub-family A, member 2 (ATP-binding cassette transporter 2) (ATP-binding cassette 2).
 GN ABCA2 OR ABC2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX PubMed=11178988;
 RA Kaminski W.E., Plehler A., Pullmann K., Porsch-Oczurne M., Duong C., Bared G.M., Buchler C., Schmitz G.;
 RT "Complete coding sequence, promoter region, and genomic structure of the human ABCA2 gene and evidence for sterol-dependent regulation in macrophages";
 RL Biochem. Biophys. Res. Commun. 281:249-258(2001)
 CC -I- FUNCTION: PROBABLE TRANSPORTER, ITS NATURAL SUBSTRATE HAS NOT BEEN FOUND YET. MAY HAVE A ROLE IN MACROPHAGE LIPID METABOLISM AND NEURAL DEVELOPMENT.

CC -I- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -I- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCA SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 DR EMBL: AF327705; AAK14335.1;
 DR EMBL: AF327658; AAK14335.1; JOINED.
 DR EMBL: AF327659; AAK14335.1; JOINED.
 DR EMBL: AF327660; AAK14335.1; JOINED.
 DR EMBL: AF327661; AAK14335.1; JOINED.
 DR EMBL: AF327662; AAK14335.1; JOINED.
 DR EMBL: AF327663; AAK14335.1; JOINED.
 DR EMBL: AF327664; AAK14335.1; JOINED.
 DR EMBL: AF327665; AAK14335.1; JOINED.
 DR EMBL: AF327666; AAK14335.1; JOINED.
 DR EMBL: AF327667; AAK14335.1; JOINED.
 DR EMBL: AF327668; AAK14335.1; JOINED.
 DR EMBL: AF327669; AAK14335.1; JOINED.
 DR EMBL: AF327670; AAK14335.1; JOINED.
 DR EMBL: AF327671; AAK14335.1; JOINED.
 DR EMBL: AF327672; AAK14335.1; JOINED.
 DR EMBL: AF327673; AAK14335.1; JOINED.
 DR EMBL: AF327674; AAK14335.1; JOINED.
 DR EMBL: AF327675; AAK14335.1; JOINED.
 DR EMBL: AF327676; AAK14335.1; JOINED.
 DR EMBL: AF327677; AAK14335.1; JOINED.
 DR EMBL: AF327678; AAK14335.1; JOINED.
 DR EMBL: AF327679; AAK14335.1; JOINED.
 DR EMBL: AF327680; AAK14335.1; JOINED.
 DR EMBL: AF327681; AAK14335.1; JOINED.
 DR EMBL: AF327682; AAK14335.1; JOINED.
 DR EMBL: AF327683; AAK14335.1; JOINED.
 DR EMBL: AF327684; AAK14335.1; JOINED.
 DR EMBL: AF327685; AAK14335.1; JOINED.
 DR EMBL: AF327686; AAK14335.1; JOINED.
 DR EMBL: AF327687; AAK14335.1; JOINED.
 DR EMBL: AF327688; AAK14335.1; JOINED.
 DR EMBL: AF327689; AAK14335.1; JOINED.
 DR EMBL: AF327690; AAK14335.1; JOINED.
 DR EMBL: AF327691; AAK14335.1; JOINED.
 DR EMBL: AF327692; AAK14335.1; JOINED.
 DR EMBL: AF327693; AAK14335.1; JOINED.
 DR EMBL: AF327694; AAK14335.1; JOINED.
 DR EMBL: AF327695; AAK14335.1; JOINED.
 DR EMBL: AF327696; AAK14335.1; JOINED.
 DR EMBL: AF327697; AAK14335.1; JOINED.
 DR EMBL: AF327698; AAK14335.1; JOINED.
 DR EMBL: AF327699; AAK14335.1; JOINED.
 DR EMBL: AF327700; AAK14335.1; JOINED.
 DR EMBL: AF327701; AAK14335.1; JOINED.
 DR EMBL: AF327702; AAK14335.1; JOINED.
 DR EMBL: AF327703; AAK14335.1; JOINED.
 DR EMBL: AF327704; AAK14335.1; JOINED.
 DR EMBL: AF327705; AAK14335.1; JOINED.
 DR EMBL: AF327706; AAK14335.1; JOINED.
 DR EMBL: AF327707; AAK14335.1; JOINED.
 DR EMBL: AF327708; AAK14335.1; JOINED.
 DR EMBL: AF327709; AAK14335.1; JOINED.
 DR EMBL: AF327710; AAK14335.1; JOINED.
 DR EMBL: AF327711; AAK14335.1; JOINED.
 DR EMBL: AF327712; AAK14335.1; JOINED.
 DR EMBL: AF327713; AAK14335.1; JOINED.
 DR EMBL: AF327714; AAK14335.1; JOINED.
 DR EMBL: AF327715; AAK14335.1; JOINED.
 DR EMBL: AF327716; AAK14335.1; JOINED.
 DR EMBL: AF327717; AAK14335.1; JOINED.
 DR EMBL: AF327718; AAK14335.1; JOINED.
 DR EMBL: AF327719; AAK14335.1; JOINED.
 DR EMBL: AF327720; AAK14335.1; JOINED.
 DR EMBL: AF327721; AAK14335.1; JOINED.
 DR EMBL: AF327722; AAK14335.1; JOINED.
 DR EMBL: AF327723; AAK14335.1; JOINED.
 DR EMBL: AF327724; AAK14335.1; JOINED.
 DR EMBL: AF327725; AAK14335.1; JOINED.
 DR EMBL: AF327726; AAK14335.1; JOINED.
 DR EMBL: AF327727; AAK14335.1; JOINED.
 DR EMBL: AF327728; AAK14335.1; JOINED.
 DR EMBL: AF327729; AAK14335.1; JOINED.
 DR EMBL: AF327730; AAK14335.1; JOINED.
 DR EMBL: AF327731; AAK14335.1; JOINED.
 DR EMBL: AF327732; AAK14335.1; JOINED.
 DR EMBL: AF327733; AAK14335.1; JOINED.
 DR EMBL: AF327734; AAK14335.1; JOINED.
 DR EMBL: AF327735; AAK14335.1; JOINED.
 DR EMBL: AF327736; AAK14335.1; JOINED.
 DR EMBL: AF327737; AAK14335.1; JOINED.
 DR EMBL: AF327738; AAK14335.1; JOINED.
 DR EMBL: AF327739; AAK14335.1; JOINED.
 DR EMBL: AF327740; AAK14335.1; JOINED.
 DR EMBL: AF327741; AAK14335.1; JOINED.
 DR EMBL: AF327742; AAK14335.1; JOINED.
 DR EMBL: AF327743; AAK14335.1; JOINED.
 DR EMBL: AF327744; AAK14335.1; JOINED.
 DR EMBL: AF327745; AAK14335.1; JOINED.
 DR EMBL: AF327746; AAK14335.1; JOINED.
 DR EMBL: AF327747; AAK14335.1; JOINED.
 DR EMBL: AF327748; AAK14335.1; JOINED.
 DR EMBL: AF327749; AAK14335.1; JOINED.
 DR EMBL: AF327750; AAK14335.1; JOINED.
 DR EMBL: AF327751; AAK14335.1; JOINED.
 DR EMBL: AF327752; AAK14335.1; JOINED.
 DR EMBL: AF327753; AAK14335.1; JOINED.
 DR EMBL: AF327754; AAK14335.1; JOINED.
 DR EMBL: AF327755; AAK14335.1; JOINED.
 DR EMBL: AF327756; AAK14335.1; JOINED.
 DR EMBL: AF327757; AAK14335.1; JOINED.
 DR EMBL: AF327758; AAK14335.1; JOINED.
 DR EMBL: AF327759; AAK14335.1; JOINED.
 DR EMBL: AF327760; AAK14335.1; JOINED.
 DR EMBL: AF327761; AAK14335.1; JOINED.
 DR EMBL: AF327762; AAK14335.1; JOINED.
 DR EMBL: AF327763; AAK14335.1; JOINED.
 DR EMBL: AF327764; AAK14335.1; JOINED.
 DR EMBL: AF327765; AAK14335.1; JOINED.
 DR EMBL: AF327766; AAK14335.1; JOINED.
 DR EMBL: AF327767; AAK14335.1; JOINED.
 DR EMBL: AF327768; AAK14335.1; JOINED.
 DR EMBL: AF327769; AAK14335.1; JOINED.
 DR EMBL: AF327770; AAK14335.1; JOINED.
 DR EMBL: AF327771; AAK14335.1; JOINED.
 DR EMBL: AF327772; AAK14335.1; JOINED.
 DR EMBL: AF327773; AAK14335.1; JOINED.
 DR EMBL: AF327774; AAK14335.1; JOINED.
 DR EMBL: AF327775; AAK14335.1; JOINED.
 DR EMBL: AF327776; AAK14335.1; JOINED.
 DR EMBL: AF327777; AAK14335.1; JOINED.
 DR EMBL: AF327778; AAK14335.1; JOINED.
 DR EMBL: AF327779; AAK14335.1; JOINED.
 DR EMBL: AF327780; AAK14335.1; JOINED.
 DR EMBL: AF327781; AAK14335.1; JOINED.
 DR EMBL: AF327782; AAK14335.1; JOINED.
 DR EMBL: AF327783; AAK14335.1; JOINED.
 DR EMBL: AF327784; AAK14335.1; JOINED.
 DR EMBL: AF327785; AAK14335.1; JOINED.
 DR EMBL: AF327786; AAK14335.1; JOINED.
 DR EMBL: AF327787; AAK14335.1; JOINED.
 DR EMBL: AF327788; AAK14335.1; JOINED.
 DR EMBL: AF327789; AAK14335.1; JOINED.
 DR EMBL: AF327790; AAK14335.1; JOINED.
 DR EMBL: AF327791; AAK14335.1; JOINED.
 DR EMBL: AF327792; AAK14335.1; JOINED.
 DR EMBL: AF327793; AAK14335.1; JOINED.
 DR EMBL: AF327794; AAK14335.1; JOINED.
 DR EMBL: AF327795; AAK14335.1; JOINED.
 DR EMBL: AF327796; AAK14335.1; JOINED.
 DR EMBL: AF327797; AAK14335.1; JOINED.
 DR EMBL: AF327798; AAK14335.1; JOINED.
 DR EMBL: AF327799; AAK14335.1; JOINED.
 DR EMBL: AF327800; AAK14335.1; JOINED.
 DR EMBL: AF327801; AAK14335.1; JOINED.
 DR EMBL: AF327802; AAK14335.1; JOINED.
 DR EMBL: AF327803; AAK14335.1; JOINED.
 DR EMBL: AF327804; AAK14335.1; JOINED.
 DR EMBL: AF327805; AAK14335.1; JOINED.
 DR EMBL: AF327806; AAK14335.1; JOINED.
 DR EMBL: AF327807; AAK14335.1; JOINED.
 DR EMBL: AF327808; AAK14335.1; JOINED.
 DR EMBL: AF327809; AAK14335.1; JOINED.
 DR EMBL: AF327810; AAK14335.1; JOINED.
 DR EMBL: AF327811; AAK14335.1; JOINED.
 DR EMBL: AF327812; AAK14335.1; JOINED.
 DR EMBL: AF327813; AAK14335.1; JOINED.
 DR EMBL: AF327814; AAK14335.1; JOINED.
 DR EMBL: AF327815; AAK14335.1; JOINED.
 DR EMBL: AF327816; AAK14335.1; JOINED.
 DR EMBL: AF327817; AAK14335.1; JOINED.
 DR EMBL: AF327818; AAK14335.1; JOINED.
 DR EMBL: AF327819; AAK14335.1; JOINED.
 DR EMBL: AF327820; AAK14335.1; JOINED.
 DR EMBL: AF327821; AAK14335.1; JOINED.
 DR EMBL: AF327822; AAK14335.1; JOINED.
 DR EMBL: AF327823; AAK14335.1; JOINED.
 DR EMBL: AF327824; AAK14335.1; JOINED.
 DR EMBL: AF327825; AAK14335.1; JOINED.
 DR EMBL: AF327826; AAK14335.1; JOINED.
 DR EMBL: AF327827; AAK14335.1; JOINED.
 DR EMBL: AF327828; AAK14335.1; JOINED.
 DR EMBL: AF327829; AAK14335.1; JOINED.
 DR EMBL: AF327830; AAK14335.1; JOINED.
 DR EMBL: AF327831; AAK14335.1; JOINED.
 DR EMBL: AF327832; AAK14335.1; JOINED.
 DR EMBL: AF327833; AAK14335.1; JOINED.
 DR EMBL: AF327834; AAK14335.1; JOINED.
 DR EMBL: AF327835; AAK14335.1; JOINED.
 DR EMBL: AF327836; AAK14335.1; JOINED.
 DR EMBL: AF327837; AAK14335.1; JOINED.
 DR EMBL: AF327838; AAK14335.1; JOINED.
 DR EMBL: AF327839; AAK14335.1; JOINED.
 DR EMBL: AF327840; AAK14335.1; JOINED.
 DR EMBL: AF327841; AAK14335.1; JOINED.
 DR EMBL: AF327842; AAK14335.1; JOINED.
 DR EMBL: AF327843; AAK14335.1; JOINED.
 DR EMBL: AF327844; AAK14335.1; JOINED.
 DR EMBL: AF327845; AAK14335.1; JOINED.
 DR EMBL: AF327846; AAK14335.1; JOINED.
 DR EMBL: AF327847; AAK14335.1; JOINED.
 DR EMBL: AF327848; AAK14335.1; JOINED.
 DR EMBL: AF327849; AAK14335.1; JOINED.
 DR EMBL: AF327850; AAK14335.1; JOINED.
 DR EMBL: AF327851; AAK14335.1; JOINED.
 DR EMBL: AF327852; AAK14335.1; JOINED.
 DR EMBL: AF327853; AAK14335.1; JOINED.
 DR EMBL: AF327854; AAK14335.1; JOINED.
 DR EMBL: AF327855; AAK14335.1; JOINED.
 DR EMBL: AF327856; AAK14335.1; JOINED.
 DR EMBL: AF327857; AAK14335.1; JOINED.
 DR EMBL: AF327858; AAK14335.1; JOINED.
 DR EMBL: AF327859; AAK14335.1; JOINED.
 DR EMBL: AF327860; AAK14335.1; JOINED.
 DR EMBL: AF327861; AAK14335.1; JOINED.
 DR EMBL: AF327862; AAK14335.1; JOINED.
 DR EMBL: AF327863; AAK14335.1; JOINED.
 DR EMBL: AF327864; AAK14335.1; JOINED.
 DR EMBL: AF327865; AAK14335.1; JOINED.
 DR EMBL: AF327866; AAK14335.1; JOINED.
 DR EMBL: AF327867; AAK14335.1; JOINED.
 DR EMBL: AF327868; AAK14335.1; JOINED.
 DR EMBL: AF327869; AAK14335.1; JOINED.
 DR EMBL: AF327870; AAK14335.1; JOINED.
 DR EMBL: AF327871; AAK14335.1; JOINED.
 DR EMBL: AF327872; AAK14335.1; JOINED.
 DR EMBL: AF327873; AAK14335.1; JOINED.
 DR EMBL: AF327874; AAK14335.1; JOINED.
 DR EMBL: AF327875; AAK14335.1; JOINED.
 DR EMBL: AF327876; AAK14335.1; JOINED.
 DR EMBL: AF327877; AAK14335.1; JOINED.
 DR EMBL: AF327878; AAK14335.1; JOINED.
 DR EMBL: AF327879; AAK14335.1; JOINED.
 DR EMBL: AF327880; AAK14335.1; JOINED.
 DR EMBL: AF327881; AAK14335.1; JOINED.
 DR EMBL: AF327882; AAK14335.1; JOINED.
 DR EMBL: AF327883; AAK14335.1; JOINED.
 DR EMBL: AF327884; AAK14335.1; JOINED.
 DR EMBL: AF327885; AAK14335.1; JOINED.
 DR EMBL: AF327886; AAK14335.1; JOINED.
 DR EMBL: AF327887; AAK14335.1; JOINED.
 DR EMBL: AF327888; AAK14335.1; JOINED.
 DR EMBL: AF327889; AAK14335.1; JOINED.
 DR EMBL: AF327890; AAK14335.1; JOINED.
 DR EMBL: AF327891; AAK14335.1; JOINED.
 DR EMBL: AF327892; AAK14335.1; JOINED.
 DR EMBL: AF327893; AAK14335.1; JOINED.
 DR EMBL: AF327894; AAK14335.1; JOINED.
 DR EMBL: AF327895; AAK14335.1; JOINED.
 DR EMBL: AF327896; AAK14335.1; JOINED.
 DR EMBL: AF327897; AAK14335.1; JOINED.
 DR EMBL: AF327898; AAK14335.1; JOINED.
 DR EMBL: AF327899; AAK14335.1; JOINED.
 DR EMBL: AF327900; AAK14335.1; JOINED.
 DR EMBL: AF327901; AAK14335.1; JOINED.
 DR EMBL: AF327902; AAK14335.1; JOINED.
 DR EMBL: AF327903; AAK14335.1; JOINED.
 DR EMBL: AF327904; AAK14335.1; JOINED.
 DR EMBL: AF327905; AAK14335.1; JOINED.
 DR EMBL: AF327906; AAK14335.1; JOINED.
 DR EMBL: AF327907; AAK14335.1; JOINED.
 DR EMBL: AF327908; AAK14335.1; JOINED.
 DR EMBL: AF327909; AAK14335.1; JOINED.
 DR EMBL: AF327910; AAK14335.1; JOINED.
 DR EMBL: AF327911; AAK14335.1; JOINED.
 DR EMBL: AF327912; AAK14335.1; JOINED.
 DR EMBL: AF327913; AAK14335.1; JOINED.
 DR EMBL: AF327914; AAK14335.1; JOINED.
 DR EMBL: AF327915; AAK14335.1; JOINED.
 DR EMBL: AF327916; AAK14335.1; JOINED.
 DR EMBL: AF327917; AAK14335.1; JOINED.
 DR EMBL: AF327918; AAK14335.1; JOINED.
 DR EMBL: AF327919; AAK14335.1; JOINED.
 DR EMBL: AF327920; AAK14335.1; JOINED.
 DR EMBL: AF327921; AAK14335.1; JOINED.
 DR EMBL: AF327922; AAK14335.1; JOINED.
 DR EMBL: AF327923; AAK14335.1; JOINED.
 DR EMBL: AF327924; AAK14335.1; JOINED.
 DR EMBL: AF327925; AAK14335.1; JOINED.
 DR EMBL: AF327926; AAK14335.1; JOINED.
 DR EMBL: AF327927; AAK14335.1; JOINED.
 DR EMBL: AF327928; AAK14335.1; JOINED.
 DR EMBL: AF327929; AAK14335.1; JOINED.
 DR EMBL: AF327930; AAK14335.1; JOINED.
 DR EMBL: AF327931; AAK14335.1; JOINED.
 DR EMBL: AF327932; AAK14335.1; JOINED.
 DR EMBL: AF327933; AAK14335.1; JOINED.
 DR EMBL: AF327934; AAK14335.1; JOINED.
 DR EMBL: AF327935; AAK14335.1; JOINED.
 DR EMBL: AF327936; AAK14335.1; JOINED.
 DR EMBL: AF327937; AAK14335.1; JOINED.
 DR EMBL: AF327938; AAK14335.1; JOINED.
 DR EMBL: AF327939; AAK14335.1; JOINED.
 DR EMBL: AF327940; AAK14335.1; JOINED.
 DR EMBL: AF327941; AAK14335.1; JOINED.
 DR EMBL: AF327942; AAK14335.1; JOINED.
 DR EMBL: AF327943; AAK14335.1; JOINED.
 DR EMBL: AF327944; AAK14335.1; JOINED.
 DR EMBL: AF327945; AAK14335.1; JOINED.
 DR EMBL: AF327946; AAK14335.1; JOINED.
 DR EMBL: AF327947; AAK14335.1; JOINED.
 DR EMBL: AF327948; AAK14335.1; JOINED.
 DR EMBL: AF327949; AAK14335.1; JOINED.
 DR EMBL: AF327950; AAK14335.1; JOINED.
 DR EMBL: AF327951; AAK14335.1; JOINED.
 DR EMBL: AF327952; AAK14335.1; JOINED.
 DR EMBL: AF327953; AAK14335.1; JOINED.
 DR EMBL: AF327954; AAK14335.1; JOINED.
 DR EMBL: AF327955; AAK14335.1; JOINED.
 DR EMBL: AF327956; AAK14335.1; JOINED.
 DR EMBL: AF327957; AAK14335.1; JOINED.
 DR EMBL: AF327958; AAK14335.1; JOINED.
 DR EMBL: AF327959; AAK14335.1; JOINED.
 DR EMBL: AF327960; AAK14335.1; JOINED.
 DR EMBL: AF327961; AAK14335.1; JOINED.
 DR EMBL: AF327962; AAK14335.1; JOINED.
 DR EMBL: AF327963; AAK14335.1; JOINED.
 DR EMBL: AF327964; AAK14335.1; JOINED.
 DR EMBL: AF327965; AAK14335.1; JOINED.
 DR EMBL: AF327966; AAK14335.1; JOINED.
 DR EMBL: AF327967; AAK14335.1; JOINED.
 DR EMBL: AF327968; AAK14335.1; JOINED.
 DR EMBL: AF327969; AAK14335.1; JOINED.
 DR EMBL: AF327970; AAK14335.1; JOINED.
 DR EMBL: AF327971; AAK14335.1; JOINED.
 DR EMBL: AF327972; AAK14335.1; JOINED.
 DR EMBL: AF327973; AAK14335.1; JOINED.
 DR EMBL: AF327974; AAK14335.1; JOINED.
 DR EMBL: AF327975; AAK14335.1; JOINED.
 DR EMBL: AF327976; AAK14335.1; JOINED.
 DR EMBL: AF327977; AAK14335.1; JOINED.
 DR EMBL: AF327978; AAK14335.1; JOINED.
 DR EMBL: AF327979; AAK14335.1; JOINED.
 DR EMBL: AF327980; AAK14335.1; JOINED.
 DR EMBL: AF327981; AAK14335.1; JOINED.
 DR EMBL: AF327982; AAK14335.1; JOINED.
 DR EMBL: AF327983; AAK14335.1; JOINED.
 DR EMBL: AF327984; AAK14335.1; JOINED.
 DR EMBL: AF327985; AAK14335.1; JOINED.
 DR EMBL: AF327986; AAK14335.1; JOINED.
 DR EMBL: AF327987; AAK14335.1; JOINED.
 DR EMBL: AF327988; AAK14335.1; JOINED.
 DR EMBL: AF327989; AAK14335.1; JOINED.
 DR EMBL: AF327990; AAK14335.1; JOINED.
 DR EMBL: AF327991; AAK14335.1; JOINED.
 DR EMBL: AF327992; AAK14335.1; JOINED.
 DR EMBL: AF327993; AAK14335.1; JOINED.
 DR EMBL: AF327994; AAK14335.1; JOINED.
 DR EMBL: AF327995; AAK14335.1; JOINED.
 DR EMBL: AF327996; AAK14335.1; JOINED.
 DR EMBL: AF327997; AAK14335.1; JOINED.
 DR EMBL: AF327998; AAK14335.1; JOINED.
 DR EMBL: AF327999; AAK14335.1; JOINED.
 DR EMBL: AF328000; AAK14335.1; JOINED.
 DR EMBL: AF328001; AAK14335.1; JOINED.
 DR EMBL: AF328002; AAK14335.1; JOINED.
 DR EMBL: AF328003; AAK14335.1; JOINED.
 DR EMBL: AF328004; AAK14335.1; JOINED.
 DR EMBL: AF328005; AAK143

FT TRANSMEM 850 872 POTENTIAL.
 FT TRANSMEM 892 914 POTENTIAL.
 FT TRANSMEM 1793 1815 POTENTIAL.
 FT TRANSMEM 1846 1865 POTENTIAL.
 FT TRANSMEM 1875 1897 POTENTIAL.
 FT TRANSMEM 1904 1926 POTENTIAL.
 FT TRANSMEM 1988 2010 POTENTIAL.
 FT TRANSMEM 1025 1032 POTENTIAL.
 FT NP_BIND 2088 2095 ATP (POTENTIAL).
 FT NP_BIND 14 14 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 90 90 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 169 169 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 174 174 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 306 306 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 369 369 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 380 380 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 421 421 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 433 433 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 477 477 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 485 485 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 495 495 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 531 531 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 545 545 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 591 591 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 601 601 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 629 629 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1409 1409 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1497 1497 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1550 1550 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1558 1558 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1613 1613 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1678 1678 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1776 1776 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2055 2055 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SO SEQUENCE 2436 AA: 269971 MM: 9E6688D8615DE0ED CRC64;

Query Match 35.9%; Score 4230.5; DB 1; Length 2436;
 Best Local Similarity 39.8%; Pred. No. 4.5e-257;
 Matches 998; Conservative 347; Mismatches 730; Indels 435; Gaps 61;

QY 6 QRLRLMKNLFRRROTCLLEVAMPLEFLLISVLSYRPEQHECHFPKA-MPSA 64
 DB 6 QLOLLMKNVTLKRRSPVLAFLFIFLVFLILGLRQKPTISYKEVPFYTAALSTA 65
 QY 65 GTLPWVOCITCANPCGRYPTPBEARGVVGNENKSIVARLFDARLL-LLYQOKDT 120
 DB 66 GILPVMOSLCPDGGORDEFGFL-QLANSTVTLLEKIDRVNEEGLFDPARP 115
 QY 121 SMKDMKRVLTLOQIKKSSNMLKLODFLVNDFSG-FLYHNLSLP 165
 DB 116 STGSELEALR-QHLEALSAGRTSGSHLDSTVSFSLDSVARNPQELWFLTONLSLP 173
 QY 166 KSTVDKMLRADY-ILAKYFLOGYQHLHS-LCNGS-KSEEMI-204
 DB 174 NSTAQLLAARVDPPEVYHLLFEPSSALDSQSGHLGQEPMSRLGNGLRFMEDILLAPA 233
 QY 205 -GQGEVSELCGLPKEXLAAB-RVLRSNMDI 235
 DB 234 LLEQLTCTGSGELGRLITVPESQKALOGYDAVCSQAARARFSGLSAELRNOLDV 293
 QY 236 LKPIRLTINSTSPFSKELAER-TKTLHSIGTIAQOLFMSKRSMDKROEVMELTVNNS 293
 DB 294 AK-VSQQLGLDAFNGSDSSPQAPPRRLDALLGLLD-AQVLODVVLIS 341
 QY 294 SSSSTQIYAVSVIYCGHPEGGLIKISLWYEDNNYKALFGNGTEDEATFEYDNSTTP 353
 DB 342 ALALLLPQHCIGRTGPPRPSAGNGAAN-GTGAAGVMPNATADGAPASAAALATP 396
 QY 354 YCNLDKLNLESPLSRIIMKALKPLLV-381
 DB 397 -DTLOGGCSATVQ-LWAGLOPILGNNRTIEPEALRGNMSSIGFTSKDQRLGLLV 451
 QY 382 -KILYEDTPATIRYMAEVNKTFOELAVFHDLEGWMEELSPKIMTFMENSQEMD 434

DB 452 HLMTSNPKIILAPAGSEVDRIILKANETFAFVGWVTHYAQVWLNISAEIRSFLEGRLOO 511
 QY 435 LVRLMLDSRDNDHFEQOOLGDMTADDTAFLAKHEDQSSNGSVYWRFAFNEN-492
 DB 512 HLKWL-OQYVAELRLPE--ALNLSIDELPRLRDNFS 547
 QY 493 -QALRTIS--RMECVNLKLEPIATEFWLJNKSME-LDERKFWAG 535
 DB 548 LPSGALLQOOLDITDNAAACGHIQMSKVSVDIFKGFPEDESIYVTLNQAQDNVTVFAS 607
 QY 536 IVFTGILTPGSLELRPHVKKIIMDIDNVERTNKTIDGQYDPGRPADFEDEMRVWGGFAV 595
 DB 608 VIFQFQKQDS-LRPHVHYKIRONSFEKTEINELIRAWRGRPMWG--GRFPLGFW 662
 QY 596 LODVVEQALIRRLTGE-KRTGVYQMOBPYCYDDIFLRYMSKMPLEMTLAMIYSVAV 654
 DB 663 IDMMERAIIDFVGDVVEPQSVQMPRYPCYTRDDPLFVIEHMMPLQWISVSVAM 722
 QY 655 IIRGIYEREARLEKEMRIMGDINSILMFSWFISSLPLVLSAGLVVILKGMLEPYSD 714
 DB 723 TIQIYAEKHEHRLKEVKTMGILNNAVHWAMFITGFVOLSIVTALITLKYGQVLMHSH 782
 QY 715 PSVYVFLSVFVNTTLOCFELISTLESRANLAACGGIIFLYLYPYLC--VAMOD 769
 DB 783 VVIIMFLAYAVATIMFCFLSVLSKAKLASACGGIIFLYLYPYLC--VAMOD 842
 QY 770 YVGFELTKIFSLPVAFGCEYFALFEEOGIGVQMDNLFESVPEDEGPNLTISIMML 829
 DB 843 ITAFK-KIASLMSITAFGLSGSKYFALYEVAGVIGIOMTFSSQPEGDDFNLLAVTLM 901
 QY 830 FDFELVGMVMTYEAVPFGQYGIIPRWYFPCIKSYWFG--ESDEKSHPSGNOKRMS--884
 DB 902 VDAVVYGLITMYEAVHPGMYGIPRPWYFPLQKSYWLGSGRLEAWKSPWAKRTRLSV 961
 QY 885 -ETC-----MEEPNHLKIGVSIQNLKVVYRDGMKVAVDGLALNFEYGOIT 929
 DB 962 EEDACAMESRREPERTEGMEEPNHLPLVVCYDPLKLVYKDKOKLALNLSINLEYNOV 1021
 QY 930 SFLGHNAGKTTMSILTLGPPTSGTAYILGKDISEKSTIRONLCVQPOHNVFDMIT 989
 DB 1022 SFLGHNAGKTTMSITLGLPPTSGTAYILGKDISEKSTIRONLCVQPOHNVFDMIT 1081
 QY 990 VEEHIMFARLKSGISGKHYVAKEMOMALDVLPSKLSKTSQSGMOKRLSVATAFVG 1049
 DB 1082 VEEHIMFARLKSGISGKHYVAKEMOMALDVLPSKLSKTSQSGMOKRLSVATAFVG 1140
 QY 1050 GSKVYILDEPTAGVDPYSRKGIWELLKYRQGTIILSTHMDADVDLRIALISHGL 1109
 DB 1141 GSRALILDEPTAGVDPYARRAILMDILKYKGRITILSTHMDADVDLRIALISHGL 1200
 QY 1110 CCVGSFLKNOGLGYLTLVKKDYESSLSGNSSTYSYLKEDSVSSSDAGLS 1169
 DB 1201 KCCGSPFLKGYGGLYTLVKKRPAEPG--GQFQGLAS 1238
 QY 1170 DHESDPTLDVS--AISNLIRKHVSEARLVEDIGHETLYVLPYEAKEGAFVLEPHIID 1227
 DB 1239 SPFGARPLSSCELOVQFIRKHVASCLVSDTSTELSYILPSEAAKGAFFLFOHLER 1298
 QY 1228 RUSDGISTSYGISETTLEIFLKAAB--SGVNA-ETSDQTL-1267
 DB 1299 SLDALHLSSFGIMDTLLEVLFLKVSDEDOSELSNSADVKEKSKVDLFGACBPASGEBHAG 1358
 QY 1268 -ARRNRA-FGDKQSCILRPFTEDDAADPNDSDIPESR 1303
 DB 1359 NLRGSELQSOASLOSASVSGARGDEAGCTVYGYRPLF-DNPDPD--NVSLOEV 1415
 QY 1304 ETDLLSGMDKGSYQYKGLTQOQFVALLKRLIARRSKRGFFAQVILPAVFPALV 1363
 DB 1416 EBAVALSRV-GQGSRLKLDGMLKVRFHGLLVKRFICARNSALFESOLLPFAFVYVAMT 1474
 QY 1364 FSLIVPFGKPSLELQPMWYNQYV--FVSNDAPE--DTGTLELLNAL 1408

Db	1475	VALSPEIGDLPPLVLSPOYH-NTTQPRGNFTPYANFERREYRLSPDASPQOLVSTF	153
Qy	1409	TKDGEFGTRCH-----EGNPI-----	1424
Db	1534	RLPSGVATGCVLSPANGSLGFTLMLSGEERLLAARFPMSCLESTGQLPLSNFVPP	1593
Qy	1425	-----P-----TPQADEEMTAP--VPQTIMDLFQNGNWTQMSPAC	1463
Db	1594	PSPASDPSASPEDDQANNVSLPRTAPREMTASPLRIIVEPR-----C	1641
Qy	1464	OCSSDKTKKMLPYCPGAGGLRPPQKONTADILQDTLGNRIDYLVKTYGQITAKSLKN	1523
Db	1642	TCSAGGTGES---CPSSVVG--HPQCMRVVGTDLITDLTGNNSEYLLTFSDRF-----	1600
Qy	1524	KIWNFEFRGGESLGVSNTQALPSPQOEYNOIAIKOMKHKLKADSCADRFNLSTGERMTG	1583
Db	1691	-----RLHRTGATITFG--NYLKSIPASFGTRAPRVRK-----	1721
Qy	1584	LDRNNVYKWFNNKGMHAIISFLVANNALIRANLQGE--NPSHGITAFLNPLMLKQ	1642
Db	1722	IAVRAAYQFYNNKQGHSMPLYLNSLNALIRANLPKSGKANAAGGITVTHNPMKNTAS	1781
Qy	1643	LSEVALMTTSVLVLSICVIFAMSFVPSFVFLIOBRYKSKAKHLOFTSGKPYVIMLSN	1702
Db	1782	LS-LDYLLQGTIVYTAIRLIIVAMSFVPSFVFLVAKESTKAKHLOFTSGCNPLITTMAN	1840
Qy	1703	FVWMDNMYVPAFTVLTITFIIFQOQKSVYSTNLPVALLLLLYGNSTPLMTPASVFKI	1762
Db	1841	IYVMDMLNLYVPAFTCVITLTFVFDLPATISPTNPAALSLELLYGNSTPIYVPAFVEFV	1900
Qy	1763	PSRTAYVLTLYNVLFTGINSVAFVLELFP--DKLNTINIDLSVFLTFPFGICRGRLD	1821
Db	1901	PSSAVYFLIVNLFTGITAIVATVFLQDLFEBHDKDKLVVNSYLKSCFLFPPVNLGHQME	1960
Qy	1822	MYKNQAMADALERGE--NREVSPLMDIVGRNLFMANBVGVFLLITVLIOYRFFIRPR	1880
Db	1961	MAYNEYINETYAKTGQDKKKSPFEMDIYTKCLVMAAEGVGVGLTLMQYNFLRRPQ	2020
Qy	1881	VNAKSLPNDEDEDVRRERORILIDGGGONDIETIKELIKITYRRK---RKAVRIGVAP	1937
Db	2021	MPVSTKPYED-DVDVASERQVRLRGADADNMYKIMNLTKVYKSKIGRIIILAVRDLGAR	2079
Qy	1938	PGECHGLLVNGACKSTFEMLTGDTITVYRGDAFLNKSLISLTHEYHOMNGYCPQDAI	1997
Db	2080	LGECGLLVNGACKSTFEMLTGDESTGGAFAVNGSHVYLKELLQYQOSTLGYCPCQAL	2139
Qy	1996	TELLTGREHVEFFALRGVPEKVEGVGEMALRTGLVKKYGEKAGNYSQGNKRKLTAM	2057
Db	2140	FDELITAREHLOLYLTRRGISWMDKARVYKMALEKLELTKIADKPAAGTYSQGNKRKLTAI	2199
Qy	2058	ALLIGRPVVELDPTTGMDPKARFLIMCALSIVYKEGSHVYLISHSMEDECACTRMAT	2117
Db	2200	ALITYPAFIFLDEPTTGMDPKARFLIMLIDLITKTSQSVYLISHSMEDECACTRLALIM	2259
Qy	2118	VNGRFRLGSVOHLKRFEGDGTIVYRILAGSNPLDKYVDFEGLAFPGSVYLKEKHRNMQ	2177
Db	2260	VNGRLCTGSIQHLKRFEDGIMITVR--TKSSQSGKVYVFFRNRPPEALMKERRHTKVQ	2318
Qy	2178	YQLEPSSLSLARLFSILSQSKRLHIEDYSVQYTTDQVFNFAKQSDN	2227
Db	2319	YQLEKSHISIAQVFSKMEQVSGVCLIEDYSVQYTTDQVFNFAKQSDN	2368


```

OY 1731 SSNTLPVALALLLLXGNSITPLMPASFVEKIPSTAYVILTSVNLFTGINSVATEVLEL 1790
DB 1869 SPTNPPAVSLFLILGWSTITPLMPASFVEKIPSTAYVILTSVNLFTGINSVATEVLEL 1928
OY 1791 PFDNKLNNIDILKSVLFPFHCIGRLIDWKNOAMDALERPEE-NREVSPLSMWL 1848
DB 1929 FHHDKLVVNSYLSKCLFIPVNLGLGEMAYNEYINERYAKIGOPKMSPEEMDI 1988
OY 1849 VGRNLFMAAVGAVFELLTVLIDYRFTIRPPVNAKLSPLNDEDEDVRRERQRIIDLGAGQ 1908
DB 1989 VTRGLVANTVAGVGFELLTMCQYNFROFORLPVSTKPEVD-DVAVASRQVNLGDAD 2047
OY 1909 NDLEIKELTKYRBR--RKPAVDRIICVIGIPCECGGLGVACKSSFFKMLTDTTV 1965
DB 2048 NDMVKIEMITKVYKSRKIGRLAVDRCLCVRGECGLGVAGAKRTSFFKMLTDEST 2107
OY 1966 TRGDAFLKNSIISNIEHVQNMKGQCFATITELLTGREHVEFALLRGVPEKEVKG 2025
DB 2108 TGGDAFVNGSHVLDLQVOOSLGYCQFPAFLDELTAHREHLOLYRLKGIIPKDEAGV 2167
OY 2026 EMATRKLGLVYKGYKGYANGVSGNKRRLSTAMALLGPPVFLDEPTGMDPKARFLMN 2085
DB 2168 KMALEKELTKYADKPAAGVYSGNKRRLSTALITGPATFLDEPTGMDPKARFLMN 2227
OY 2086 CALSVVEGSGVYLSHSMCECALCTRMALVNGRFLGVSQHLKNRFGDGYTVIRI 2145
DB 2228 LILDLITGSHVYLSHSMCECALCTRLAIWNGRFLGVSQHLKNRFGDGYTVIRI 2286
OY 2146 AGSNPDLKPYODFFGLAFGSVLEKERNMLQYOLPSSLSLARFFSLISQSKKLHED 2205
DB 2287 TKSSQVWKDVVPFNNRPFEMALKEHHTKKYOLKSEHISLAQVSKMEGVVGLIED 2346
OY 2206 YVSQTTLLQVFNFAKDSOD 2227
DB 2347 YVSQTTLLQVFNFAKDSOD 2368

RESULT 6
ABC3_HUMAN
ID ABC3_HUMAN STANDARD: PRT: 1704 AA.
AC 099758: 092473:
DT 30-MAY-2000 (Rel. 39, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE ATP-binding cassette, sub-family A, member 3 (ATP-binding cassette transporter 3) (ATP-binding cassette 3) (ABC-C transporter).
GN ABC3 OR ABC3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI-TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Thyroid carcinoma;
RX MEDLINE=96326608; PubMed=8706931;
RA Klugbauer N., Hofmann F.;
RT "Primary structure of a novel ABC transporter with a chromosomal localization on the band encoding the multidrug resistance-associated protein";
RL FEBS Lett. 391:61-65(1996).
RP [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=9719225; PubMed=9027511;
RA Connors T.D., Van Raay T.J., Petry L.R., Klinger K.W., Landes G.M., Burn T.C.;
RT "The cloning of a human ABC gene (ABC3) mapping to chromosome 1p13.3";
RC Genomics 39:231-234(1997).
CC -1- FUNCTION: MAY BE A TRANSPORTER, ITS NATURAL SUBSTRATE HAS NOT BEEN FOUND YET (BY SIMILARITY). MAY ACT AS AN EFFLUX PUMP FOR CHEMOTHERAPEUTIC DRUGS.
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN LUNG, FOLLOWED BY BRAIN,

```

```

CC PANGREAS, SKELETAL MUSCLE AND HEART. WEAKLY EXPRESSED IN PLACENTA, KIDNEY AND LIVER. ALSO EXPRESSED IN MEDULLARY THYROID CARCINOMA, CELLS (MTC) AND IN C-CELL CARCINOMA.
CC -1- DOMAIN: MULTIFUNCTIONAL POLYPEPTIDE WITH TWO HOMOLOGOUS HALVES, EACH CONTAINING AN HYDROPHOBIC MEMBRANE-ANCHORING DOMAIN AND AN -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCA SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U78735; AAC0967.1; -
DR EMBL: X97187; CA65825.1; -
DR Genew: HGNC:33; ABCA3.
DR MIM: 601615; -
DR InterPro: IPR003593; AAA_Atpase.
DR InterPro: IPR003439; ABC_transportr.
DR Pfam: PF00005; ABC_tran.2.
DR ProDom: PD000006; ABC_transportr; 2.
DR SMART: SM00382; AAA; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
DR KMW: ATP-binding; Transport; Transmembrane.
FT TRANSMEM 22 42 POTENTIAL.
FT TRANSMEM 249 269 POTENTIAL.
FT TRANSMEM 307 327 POTENTIAL.
FT TRANSMEM 344 364 POTENTIAL.
FT TRANSMEM 373 393 POTENTIAL.
FT TRANSMEM 405 425 POTENTIAL.
FT TRANSMEM 447 467 POTENTIAL.
FT TRANSMEM 925 945 POTENTIAL.
FT TRANSMEM 1100 1120 POTENTIAL.
FT TRANSMEM 1144 1164 POTENTIAL.
FT TRANSMEM 1183 1203 POTENTIAL.
FT TRANSMEM 1213 1233 POTENTIAL.
FT TRANSMEM 1245 1265 POTENTIAL.
FT TRANSMEM 1306 1326 POTENTIAL.
FT NP_BIND 566 573 ATP (POTENTIAL).
FT NP_BIND 1416 1423 ATP (POTENTIAL).
FT CONFLICT 36 36 P -> S (IN REF. 2).
FT CONFLICT 196 196 L -> P (IN REF. 2).
SQ SEQUENCE 1704 AA: 191387 MW: 46098DAF7A04F5F CRC64;

Query Match 22.4%; Score 2642.5; DB 1; Length 1704;
Best Local Similarity 35.7%; Pred. No. 1.7e-157;
Matches 639; Conservative 300; Mismatches 523; Indels 327; Gaps 48;

OY 534 AGIYFTGILPGSIE-LPHNKKYKIRMDINVERTN-----KIKDGYW----- 574
DB 131 AAVFEPNPKSKEDPLAKYHLRF--SYRRNRMWTQTSFFLKECTECHNITSLRPL 187
OY 575 --DPRADPFED--MRVVGFAVLDVVRQALIRVLGTGE-----KTIQYVQDMRY 624
DB 188 FPNPGRPEILTSPDGGEGYIRGEFLAVQNAVDKALMEYNADATQOLFQRLTVITKRPY 247
OY 625 PCYVDLFLRVMSRMPLEMTLAWISVAVIIGIYVEKEARKLTKRMINGLSDNLSLWFS 684
DB 248 PFIADPFLVAYOYOLPLLLLSFTYTALTIRAVVOEKERKRYKMYMGMGLSLWWSA 307
OY 685 WFISSLPLVSAAGLVLLIKLG-----NLIPYSDPSVVFVFLSFVAVVTLQCFILSTL 739
DB 308 WFLFLFLILIASFWTLFLCYKVPNVAVLSRSDPSLVLLFLCFAISTISFSFWSTF 367
OY 740 FSRANIAAACGGITFTTLVLYVLCVAMODYVGFLLKIFASLSPVAFGFCCEYFALFEE 799
DB 368 FSRANIAAAGFVLFYFTYIPYFVAPRYNMNLTQKSCSLLSNVAMAGAOILGKFEA 427
OY 800 OGIGVQMDLFFSSPVE-EDGFNLTTISIMMLFDPLVGVMTWYIEAVPQOYGIPTRWYF 858

```


Db 428 KGMGIOWRDL-SPVNVDDPCFGVGLMLLDSVLGLVITWMEAVFGQVGPQWYF 486
 QY PCTKSTWGE-----ESDEKSHSNOKRMSELICMEEBEPHLKLGVSINLVKVEDGM 912
 Db 487 FIMPSWCKPRAVAGKEEDSDP--EKALRNEYFEABPDVAGIKIKHLSVFRVGN 543
 QY 913 K--VAVDGLALFNYEGQITSPFLGNGAGKTTTMSLTGLFPPTSGTAVYLLGKDIRSEMT 970
 Db 544 KDBAAVRDNLMLYEGQITVLLGNGAGKTTTMSLTGLFPPTSGTAVYLLGKDIRSEMT 603
 QY 971 IRONLGVCPQHNVLFDMLTVEEHLWFTYARLKGSLSEKHYKAEEMQALDGLPSSKLSKT 1030
 Db 604 IRKSLGLCPQHILFDNLVLAHELYFAQLKGLSRQKPEEVKQMLHTLGL-EDKWNRS 662
 QY 1031 SOLSGMORKSLVALAFVGSKVVLDEPTAGVDYSPRRGIWELLKRYGRTTILSTTH 1090
 Db 663 RFLSGGRRRLSTGLALINGSKVLLDEPTSGMDAISRAITMDLQKSDRTIYLTTHF 722
 QY 1091 MDEADVLCGRALIIHSHGLCCVSSSLFLKNQIGTYLTLVKKDYESSLSGCRSSSTVS 1150
 Db 723 MDEADLLGRALIMAKGELQCCSSFLTKQYAGYHMTLVE-----P 766
 QY 1151 YLKEDSVSSSSDAGLGSHESDTLTIDVSAISNLIRKHVEARLVEDIGHELTYVLPY 1210
 Db 767 HCNED-----ISOLVHHHVFNATLLESAGALSFLIPR 800
 QY 1211 EAKKEGAFVLEPHIDRLSDLGSSYISSETTLEIEPLKVAE-ESGVDAETSDGTLPA 1268
 Db 801 ESTHR--FEGLFPAKLEKKEKELGASFGASITMEVPLRVGLVDSMDIOAIO--LPA 856
 QY 1269 ---RRNRAGF--DKOSCLPTEEDDADPND---SDIDPERFEDLLSGMDKGSVGV 1319
 Db 857 LOYOHERRASMAVDSMLC-----GAMPDSGICALIEETTAVKLWTLG----- 901
 QY 1320 KGMKLTQOQFALLMKRLLIARRSRKGFPAOIVLPAVFCIALVFSLIAPPGKYPSEL 1379
 Db 902 ---ALHCOQEFAMFLKKAASWREMKVAAQVLPVLTCTVLLAL----- 942
 QY 1380 QPMWYNEOYTFVNSDAREDTGTELLNALTGRPGFTGMEGNPFDPTGQAGEEWTTA 1439
 Db 943 -----AINYSELFDPPML--RLTLG-----EYGR 966
 QY 1440 PVPOTIMDLFONGWMTQONSPRACQSSDKIKMLPVCPPGAGGLPPQKONTADLID 1499
 Db 967 VVPSVPGTSGOOLSEHLKDALQAGS-----QPREVLID 1003
 QY 1500 LTRNISDYLVKTYVQIILAKSLKNKIWNERRY---GGSFLGVSNTQALPSEQEVNDAL 1555
 Db 1004 L-----EEFLI-----FRASVEGGGFN----- 1020
 QY 1556 KOMKHLKLAKSSADRFNLISLGRPMGTGRNNVYKWFNNKGMHAISFLVYNNALIR 1615
 Db 1021 -----ERCL--VAASFVDGERTVYVNALENNQAHSPATLAAVYDNLIFK 1063
 QY 1616 ANLQGENPNSHYGITAFNHP-----LNTLQOLSEVALMTSVDLVLCIVFAMSEVPA 1670
 Db 1064 ---LLCG---PHASIVVSNFPORSLQAKQOFNE---GRGPDIALNL--LTFMARLAS 1113
 QY 1671 SEVVELLQERVSRAKHLQFISGVKPVYIWSLNFWDKMCNYYVPTVYLLIFTCQOKSYV 1730
 Db 1114 TFSILAVSRVAQAHVQSVGSVHASFWSLALLMDLISFLIPSLILLVYKADVRAFT 1173
 QY 1731 SSTNLPVALALLLQGMSTTLMYSPASVFKIPSTAYVVLVSVMLFGIGINSVTFVLEL 1790
 Db 1174 RQGHMADTLLLLLGWALITPLMYLMNEFFLGAATAYTRLLITFNLSI-----ATFLMVT 1229
 QY 1791 ---FTDNKLNINDILKSVFLFPHFCGLRGLIDWYKN-----QAMADALERP 1836
 Db 1230 IMIIPAVKLEELSKLDHVFVLPHNCLGMAVSEYENYERFRYCTSESEVAHCKKXNI 1289
 QY 1837 ---ENRNPVLSMDL--VGRNLFMAVSEVVELTIVLIOYRFFIRPPVANKL----- 1885
 Db 1290 OYOENFY---AWSAPGVGRFVASMASGCAVLLIFLETINLQRLGLICALRRKRL 1345

QY 1886 -----SPUNDEDEYRRERORIIDGGGNDI---LEIKELTKYRRKKRP--AVDRICV 1934
 Db 1346 TELYTRMPELDQDVADERTRIILASPDSLLRPLIKELSKYV--QQRPLLAVALRSL 1404
 QY 1935 GTRPCEFCGLLVNGAGKSSFEKMLTGDTTVYTRGDAFLNKSILSNIHEVQNMKCPQF 1994
 Db 1405 AVQKCEGCLLOFNAGKTTTFKMLTGCEESLTSGDAFVGHRISSDGKVRORIGYCPQF 1464
 QY 1995 DATTELTGREHVEFALLRGVPEKEVGKVEDMAIRKIGLVKYGEKXAGVNSGCKRKL 2054
 Db 1465 DALDHMTGRELVMYARLGRIPERHIGACVENTLRLGLLEPHANKLVRYTSGCKRKL 1524
 QY 2055 TAMALIGPVPVFLDEPTGMDPKARREFLMNCALSVVEGSRVLTSHSMECEALCTRM 2114
 Db 1525 TGALIGEPVAVFLDEPTGMDPKARREFLMNCALSVVEGSRVLTSHSMECEALCTRL 1584
 QY 2115 AIMVNGRFRCLGSVOHLKKNRFGDYTYTVRT--AGSNPLKPVODFGLAFPGSVLKEKH 2172
 Db 1585 AIMVQGFCLGSPQHLKSKFGSGYSLRAKVQSGQOALEEFAFYDLTPPGSVLEDEH 1644
 QY 2173 RNMLOYQLPSSLSLARIFSLISQSKRLHIEDYSVSTQTPDVPVNFPA 2221
 Db 1645 QGMVHYHLPGRLDSMAKVFGLLEKAKETGVDDYSQISLEOVFLSFA 1693

RESULT 7
 CED7_GAEL STANDARD: PRT: 1704 AA.
 ID CED7_GAEL P34359; 076287;
 AC P34358; P34359; 076287;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE ABC transporter ced-7 (Cell death protein 7).
 GN CED-7 OR C48B4.4.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 OX NCBI_Taxid=6239;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM C), FUNCTION, AND MUTAGENESIS OF LYS-586;
 RP GLO-639 AND LYS-1417.
 RC STRAIN-Bristol N2;
 RC MEDLINE=98297348; PubMed=9635425;
 RA Wu Y.-C., Horvitz H.R.;
 RT "The C. elegans cell corpse engulfment gene ced-7 encodes a protein
 RL similar to ABC transporters.";
 RL Cell 93:951-960(1998).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Bristol N2;
 RC MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Furlon M., Gader A., Green P., Hawkins T., Hillier L., Jier M.,
 RA Johnston L., Jones M., Kershaw J., Kisten J., Laister N.,
 RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopia A., Saunders D., Showkhen R.,
 RA Sims M., Smaildon N., Smith A., Smith M., Sonhammer E., Staden R.,
 RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
 RA Watson R., Watson A., Winstock L., Wilkinson-Sproat J.,
 RA Wohldman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RL elegans.";
 RL Nature 368:32-38(1994).
 [3]
 RP REVISIONS, AND ALTERNATIVE SPLICING.
 RA Durbin R.;
 RL Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.
 CC -i- FUNCTION: Functions in the engulfment of cell corpses during
 CC embryonic programmed cell death to translocate molecules that
 CC mediate homotypic adhesion between cell surfaces of the dying and

CC engulfing cells.
 CC - SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC - ALTERNATIVE PRODUCTS: 3 isoforms, a, b and c (shown here); may be
 CC produced by alternative splicing.
 CC - TISSUE SPECIFICITY: Ubiquitous in embryos. Expressed in larval
 CC germ-line precursors. Expression in larvae and adults is seen in
 CC sheath cells, pharyngeal-intestinal valve and phanid
 CC sheath cells. Low levels of expression are also seen in gonadal
 CC sheath cells.
 CC - DOMAIN: Multifunctional polypeptide with two homologous halves,
 CC each containing a hydrophobic membrane-anchoring domain and an
 CC ATP binding cassette (ABC) domain.
 CC - SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCA SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF049142; AAC24116.1; -
 CC EMBL; Z29117; CA82384.2; -
 CC EMBL; Z29117; CA82383.2; -
 CC EMBL; Z29117; CAC4271.1; -
 CC PIR; S40724; S40724.
 CC PIR; S40725; S40725.
 CC WormPep; C48B4.4a; CE24856.
 CC WormPep; C48B4.4b; CE24857.
 CC WormPep; C48B4.4c; CE27867.
 CC InterPro: IPR003593; AAA_Atpase.
 CC InterPro: IPR003439; ABC_transport.
 CC Pfam: PF000006; ABC_tran; 2.
 CC ProDom: PD000006; ABC_transport; 2.
 CC SMART; SM00382; AAA; 2.
 CC DR POSTITE; PS00211; ABC_TRANSPORTER; 2.
 CC DR ATP-binding; Transport; Transmembrane; Repeat; Glycoprotein;
 CC KW Alternative splicing.
 CC
 CC TRANSMEM 24 44
 CC TRANSMEM 256 276 POTENTIAL.
 CC TRANSMEM 306 326 POTENTIAL.
 CC TRANSMEM 335 355 POTENTIAL.
 CC TRANSMEM 436 456 POTENTIAL.
 CC TRANSMEM 963 983 POTENTIAL.
 CC TRANSMEM 1126 1146 POTENTIAL.
 CC TRANSMEM 1176 1196 POTENTIAL.
 CC TRANSMEM 1201 1221 POTENTIAL.
 CC TRANSMEM 1234 1254 POTENTIAL.
 CC TRANSMEM 1311 1331 POTENTIAL.
 CC NE_BIND 580 587 ATP (POTENTIAL).
 CC NE_BIND 1411 1418 ATP (POTENTIAL).
 CC CARBOHYD 126 126 N-LINKED (GLCNAC).
 CC CARBOHYD 145 145 N-LINKED (GLCNAC).
 CC CARBOHYD 359 359 N-LINKED (GLCNAC).
 CC CARBOHYD 421 421 N-LINKED (GLCNAC).
 CC CARBOHYD 427 427 N-LINKED (GLCNAC).
 CC CARBOHYD 481 481 N-LINKED (GLCNAC).
 CC CARBOHYD 678 678 N-LINKED (GLCNAC).
 CC CARBOHYD 727 727 N-LINKED (GLCNAC).
 CC CARBOHYD 899 899 N-LINKED (GLCNAC).
 CC CARBOHYD 986 986 N-LINKED (GLCNAC).
 CC CARBOHYD 1012 1012 N-LINKED (GLCNAC).
 CC CARBOHYD 1045 1045 N-LINKED (GLCNAC).
 CC CARBOHYD 1597 1597 N-LINKED (GLCNAC).
 CC CARBOHYD 1632 1632 N-LINKED (GLCNAC).
 CC CARBOHYD 992 993 MISSING (IN ISOFORM A).
 CC VARSPLIC 496 508 MISSING (IN ISOFORM B).
 CC VARSPLIC 586 586 MISSING (IN ISOFORM B).
 CC MUTAGEN 586 586 K->R: CELL CORPSES NOT ENGULFED.
 CC MUTAGEN 639 639 E->G: CELL CORPSES NOT ENGULFED.
 CC MUTAGEN 1417 1417 K->R: SOME CELL CORPSES NOT ENGULFED.
 CC SEQUENCE 1704 AA; 191411 MW; B7502A0B24507CEE CRC64;

Query Match 13.0%; Score 1538.5; DB 1; Length 1704;
 Best Local Similarity 24.9%; Pred. No. 43e-88;
 Matches 484; Conservative 347; Mismatches 675; Indels 441; Gaps 59;
 458 WTAQIV-----AFLAKHPEDVSSNGSV-----TWREAFNETN-----492
 24 WLFELLIPCLLGLPLVLYLVKKNADHTSPENIYDNFGVGEVDVFLSENFKPIYKKR 83
 493 -----QAFRI-----SRFMECVNLNKL-----EPLATEVWLINKS 523
 84 CLRSQVYGYTSKRAAKRIVYDLMKFAEFQSAKIKLSKNESSEEDILT-----VLARD 140
 534 MELIDE-----RKFWAGIEFTGILPGSTEDLPHVKKYKIRMDIONVPRTKIKIDGVNDP 578
 141 LPMINEFCAINSYAAGVFEDEVDTNKKL-----NRIILGKRPETWILTESTNPGP 196
 579 RADFEDM-----RYWNGFAYLDVDVQALIRVLGTGTEKKTYVMOOMPYPYVD-----629
 197 SSGKRSRIPSSPPYWSAFILFQHAIESSFLSSVSGADPLPTTLGLPREPKTSVSA 256
 630 --DIFLRVSKMPLFMTLMIYSVAIIKGYEKEARLKEKTRIMGINSILMFSPFI 687
 257 FIDFF-----PFIWAFVTFINVIHITREIAENHA-VKPYLRAMGLSTMFYAAHV 307
 688 SSLIPLYSAGLVYIILKGNLLPYDSPVVEFVSFAVNTILOCFLISTLFSRANLAA 747
 308 MAFLEFFVI--FLCSIIPIFVMEFVSPALIVTYLMYGLGAVIFGAFVAFSPFNNTSAI 365
 748 ACGGIIFTLPLVYLCAVAMODVGFCLKIFASL--ISPV-----AGFGCEYVAL 796
 366 K-----ALIVMGMAMIGISTYLRBELDQISSCLYGLINGAPALAVEAISD 412
 797 FEEDGIGVOMDNLF-ESPVEEDGFNLTTSISMLFDTFLGYMYTIBAV-FPGQYGI 854
 413 YMRERELNLTNFMNDSSSH--FSLGVALVMIVIDILMMSIGALVVDHIFTSADFILT 469
 855 PWVFPCPKSYWFEESDEKSHNG-----SNOKRSEICMEDEP---THL-----895
 470 LFDFF-----APEDEMDGTGVTATONTRINEOVNRNRSDEIOMNPASTSLAPPN 522
 896 -----KLGVSIONLVKYYRDMGVAAVDGLALNFBEGQISPLGN 935
 523 ADSDSLEGSTADGARDADARADIIVRNLYKWTSTGSRADVGLSLRAVKGQCSILLGNH 582
 936 GAGKTTNLLGLFPPTSGTAVIIGKDIRSBASTIRONLCVCPQOHVLEFDMITVEHIV 995
 583 GAGKSTTESSTIGILPTNGRTITGIDVGNBPETRRHIGMCPQYNPDIYDLVSEHLK 642
 996 FYARLKLGLSEKHVKAEMQMALDVGLPSSKLSKTSOLSGMORKLSVALAFVGGSKYVI 1055
 643 LVYGLKAGAREKFKQDMKRILSDVKL-DFKENKAVNLGSKMKRKLCCVALLIGDESVYL 701
 1056 LDEPTAVDVPYRSRQIWEILLKYRGRTIILSTHMDADVLDGRRIAIIISHGLCCVGS 1115
 702 LDEPTAGMDPGARQOVOKLIVERKANRPTIILTYHDEARLGDWVFIHSGVLVASGIN 761
 1116 LFLKNQLTGYVILYKVDVESSLSGCRNSSSTYSYLKKEDSVQSSDAGLGSDESP 1175
 762 GYLKQKFGTGLLVY-----LDHNGDK 784
 1176 LTIYSAISNLTIRKXVSEARLVEDIGHLETVYLPYEAARKGAEVLFHEID-----1226
 785 RKMAY-ILTDVCTHYVKEAEERHGGQIETILPEARKKE--FVPLFQALEAIQDNINYS 841
 1227 -----DRISDGLISSYGSITTELEIFLKAESGVDAETSDGLPARRRRAAF 1275
 842 NVFDMNPITLKSQATLMEKSGFSLNLTLEOVETITGDK-----VDRKIASRONSRTS 894
 1276 GDKOSCLRPFTEDDAADPNDSIDIPESRETDLLSGMDGSGSYOVKWKILQOOFAVALLMK 1335
 895 HNSRNASPESLKPAQGYDTQSGTSKADSQYK--LMSDAQRP-EKSGVAMVAQOFTSIMKK 951
 1336 RLILARRSKRGFFAQVILPAVFCIALVESLIVPPRGKYPSLEIOPWMIINQYTFVSDA 1395

ID NODI_RHISN STANDARD: PRT: 343 AA.
AC P55476;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nodulation ATP-binding protein I.
GN NODI OR Y4HF.
OS Rhizobium sp. (strain NGR234).
OC Plasmid sym pNGR234a.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=394;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97305956; PubMed=9163424;
RA Freilberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,
RA Perret X.;
RT "Molecular basis of symbiosis between Rhizobium and legumes.";
RL Nature 387:394-401(1997)
CC -1- FUNCTION: FORMS, WITH NODJ, A MEMBRANE TRANSPORT COMPLEX INVOLVED
CC IN THE NODULATION PROCESS. IT PROBABLY EXPORT A MODIFIED BETA-1,4-
CC LINKED N-ACETYLGLUCOSAMINE OLIGOSACCHARIDE.
CC -1- SUBCELLULAR LOCATION: Membrane-associated.
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

DR EMBL: AE000076; AAB91694.1;
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR003439; ABC_Transport.
DR Pfam: PF00005; ABC_tran; 1.
DR ProDom: PD000006; ABC_transport; 1.
DR SMART: SM00382; AAA; 1.
DR TIGRfams: TIGR01288; nodJ; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
DR Modulation: Membrane; ATP-binding; Transport; Plasmid.
KW NP_BIND 77
FT NP_BIND 77
SQ SEQUENCE 343 AA; 37917 MW; F49AVEC56E099A33 CRC64;

Query Match 3.2%; Score 382.5; DB 1; Length 343;
Best Local Similarity 35.1%; Pred. No. 1.3e-16;
Matches 91; Conservative 46; Mismatches 101; Indels 21; Gaps 6;

QY 872 EKSHGSGN-----QKRMSEICMEEPHHLKLGVSIGNLVKVVYRDGKKAVALDGLALNFEYEG 926
DB 24 QKCHSHADNSLSRSKSDVAIE-----LTVNSKSYGD-KYVVDQLSFTITSG 70
QY 927 QITSEFLGNGAGKTTTMSITLGLFPPTSGTAVITLGDIREMSTIQLNGVCPQHNVLFD 986
DB 71 ECFGLLPNGAGKSTVSLVGLAPDEGTTITVLGEVVARARALARSRGVVPQDPFTDR 130
QY 987 MLTVEEHLPYARLKGISEKHVAKEFMQALDVGRLPSKLSKTSQSLSGMOKRLISVALA 1046
DB 131 EFTAREMLVFGKRFGLHRELEAIPPL-LDFARLESKADVPVAVQLSGMGORRLTLACA 189
QY 1047 FVGGSKVYVILDEPTAGDVPYSRGIVELLK-YRGRITITLSTHMDADYLDGRIATIS 1105
DB 190 LINDPOLLLIDEPDTGDPHARHLINERLSLALGKLTLLTHHEMRADRLDCLVYIE 249
QY 1106 HGLTCCVG-SLFLKQNLG 1123
DB 250 HGKRIYGRPHALIDEOIG 268

RESULT 10
NODI_RHIS3
ID NODI_RHIS3 STANDARD: PRT: 304 AA.

AC P72335;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nodulation ATP-binding protein I.
GN NODI.
OS Rhizobium sp. (strain N33).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=103798;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96303537; PubMed=8755627;
RA Cloutier J., Laberge S., Prevost D., Antoun H.;
RT "Sequence and mutational analysis of the common nodBCJ region of
RT Rhizobium sp. (Oxytropis arctobia) strain N33, a nitrogen-fixing
RT microsybiont of both arctic and temperate legumes.";
RL Mol. Plant Microbe Interact. 9:523-531(1996).

CC -1- FUNCTION: FORMS, WITH NODJ, A MEMBRANE TRANSPORT COMPLEX INVOLVED
CC IN THE NODULATION PROCESS. IT PROBABLY EXPORT A MODIFIED BETA-1,4-
CC LINKED N-ACETYLGLUCOSAMINE OLIGOSACCHARIDE.
CC -1- SUBCELLULAR LOCATION: Membrane-associated.
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

DR EMBL: U53327; AAB16998.1;
DR HSSP: O58663; 166H.
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR003439; ABC_Transport.
DR Pfam: PF00005; ABC_tran; 1.
DR ProDom: PD000006; ABC_transport; 1.
DR SMART: SM00382; AAA; 1.
DR TIGRfams: TIGR01288; nodJ; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
DR Modulation: Membrane; ATP-binding; Transport.
KW NP_BIND 38
FT NP_BIND 38
SQ SEQUENCE 304 AA; 33698 MW; 7C6A33B0364CCE14 CRC64;

Query Match 3.1%; Score 366; DB 1; Length 304;
Best Local Similarity 35.0%; Pred. No. 1.2e-15;
Matches 77; Conservative 54; Mismatches 85; Indels 4; Gaps 3;

QY 896 KLGVSIGNLVKVVYRDGKKAVALDGLALNFEYEGQITSEFLGNGAGKTTTMSITLGLFPPTSG 955
DB 3 KVAIDLAVGKRSFGD-KLVVNLSTFVASGEFGTLGNGAGKSTIARMLLGMVPPAG 60
QY 956 TAYITLGDIREMSTIQLNGVCPQHNVLFDMLTVEEHLPYARLKGISEKHVAKEFMQALDVGRLPSKLSKTSQSLSGMOKRLISVALA 1015
DB 61 KITVYLPYGAARSLARKISGVVPQDPNDQEFVRENILVGFVSGSTKIK-EVIPS 119
QY 1016 ALDVGLPSSKLSKTSQSLSGMOKRLISVALAVGSKVYVILDEPTAGDVPYSRGIVELLK-YRGRITITLSTHMDADYLDGRIATIS 1074
DB 120 LLEFARLESKADVPVAVQLSGMGORRLTLAKALLINDPOLLVDEPTGDPHARHLINERL 179
QY 1075 LKVRGRITITLSTHMDADYLDGRIATISHGKLCVGS 1114
DB 180 RFLLAGKTIITLTHHEMRADRLDCLVLEHGKRLAGS 219

RESULT 11
YADG_ECOLI
ID YADG_ECOLI STANDARD: PRT: 308 AA.
AC P36879;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Hypothetical ABC transporter ATP-binding protein yadG.
GN YADG OR B0127.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN-K12 / W3110;
RC MEDLINE=94261430; PubMed=8202364;
RA Fujita N., Mori H., Yura T., Ishihama A.;
RT "Systematic sequencing of the Escherichia coli genome: analysis of
RT the 2.4-4.1 min (110,917-193,643 bp) region.";
RL Nucleic Acids Res. 22:1637-1639(1994).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RC MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Blattner F.R., Plunkett G. III, Glasner J.D., Rode C.K., Mayhew G.F.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1232-1244(1997).
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. STRONG, TO
CC RHIZOBIACEA NODI.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D26562; BAA05583.1;
DR EMBL: AE000122; AAC73238.1;
DR PIR: S45204; S45204.
DR Ecogen: Egi2320; yadG.
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR003439; ABC_Transport.
DR Pfam: PF00005; ABC_tran. 1.
DR Prodom: PD000006; ABC_transport. 1.
DR SMART: SM00382; AAA. 1.
DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
DR Hypothetical protein; ATP-binding; Transport; Complete proteome.
KW NP_BIND 38 45 ATP (POTENTIAL).
FT SEQUENCE 308 AA; 34647 MW; 9EFA3C2BFFBF3E47 CRC64;
SQ
Query Match 2.9%; Score 347; DB 1; Length 308;
Best Local Similarity 28.2%; Pred. No. 1.9e-14;
Matches 104; Conservative 64; Mismatches 133; Indels 68; Gaps 8;
QY 895 LKLGSIQNLVYVYRDGKAVAVDGLALNFYEGQITSFGLHNGAGKTTYSILGLPPTS 954
DB 1 MTALEIQDLKTYGQV-ALRGIDLOVEAGDEYALLGPGACKSTYTGIISSLVKTS 59
QY 955 GTAYILGKDIREKSTIQNLGVCQGHVLFDMLEVEHIFVYARLGLSEK--HVAEM 1012
DB 60 GRVAFVGYDLEKDVNARQQLDLPQEFNFNPFETVQIVVNOAGYGVKEKAVITSEK 119
QY 1013 EQMALDVGIPSSKLSKTSQSLGQMRKLSVALAFVGGSKVILDEPTAGVDPYSRGIW 1072
DB 120 YLKQDLW---GKRNRRAVMSLGGKMRRLMTARALMHPKLLIIDEPAVDIELRSMW 176
QY 1073 ELLKTR-OGRTIIISTHMDADVGLRIALISHCXLCGVSSSLFLKNQGTGYTLV 1131
DB 177 GFLKLDNDGTTILITLTHLEAEMLCRNIGITGHELV----- 215
QY 1132 KRVESLSGSRSSSTVYLKKEKDSVSQSSDAGLGDSDHDTLTIDVSAISLIRKHV 1191
DB 216 -----ENTSMKALLAKLSEFTLIDLAFLAPSLPKIDG 247

QY 1192 SEARLVEDIGHELTIVLPYEAKGAFVELFHEIDRLSDIGS--SYGSEPTLEELF 1249
DB 248 YQYRLVD-----TATLEVEVLRQGINSVF-----TQLSQGIQVLSMKNKRRLEELV 297
QY 1250 KVAESGVD 1258
DB 298 SILVERKQCD 306
RESULT 12
NODI_RHILIO
ID NODI_RHILIO STANDARD: PRT: 340 AA.
AC P23703;
DT 01-NOV-1991 (Rel. 20, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Nodulation ATP-binding protein I.
GN NODI OR MLR6164.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN-NZP 2213;
RC MEDLINE=91067466; PubMed=2251131;
RX Young C.A., Collins-Emerison J.M., Scott D.B.;
RT "Nucleotide sequence of Rhizobium loti nodI.";
RL Nucleic Acids Res. 18:6651-6691(1990).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN-MAFF303099;
RC MEDLINE=21082930; PubMed=11214968;
RX Keneo T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
CC -1- FUNCTION: FORMS, WITH NODJ, A MEMBRANE TRANSPORT COMPLEX INVOLVED
CC IN THE NODULATION PROCESS. IT PROBABLY EXPORT A MODIFIED BETA-1,4-
CC LINKED N-ACETYLGLUCOSAMINE OLIGOSACCHARIDE.
CC -1- SUBCELLULAR LOCATION: Membrane-associated.
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X55705; CAA39236.1; ALT_INIT.
DR EMBL: AP003008; BAB52501.1;
DR PIR: S13590; S13590.
DR HSP: Q58663; 166H.
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR003439; ABC_Transport.
DR Pfam: PF00005; ABC_tran. 1.
DR Prodom: PD000006; ABC_transport. 1.
DR SMART: SM00382; AAA. 1.
DR TIGR: TIGR01288; nodI. 1.
DR Nodulation: Membrane; ATP-binding; Transport; Complete proteome.
KW NP_BIND 74 81 ATP (BY SIMILARITY).
FT CONFLICT 10 10 D -> E (IN REF. 1).
FT CONFLICT 97 97 T -> A (IN REF. 1).
FT CONFLICT 129 129 F -> L (IN REF. 1).
FT CONFLICT 167 167 D -> N (IN REF. 1).
FT SEQUENCE 340 AA; 37428 MW; 5777722B28D130EA CRC64;
SQ

Query Match 2.9%; Score 344.5; DB 1; Length 340;
 Best Local Similarity 32.9%; Pred. No. 3.2e-14;
 Matches 80; Conservative 51; Mismatches 93; Indels 19; Gaps 4;

OY 867 GEESEKHPGSSNOKRMSKELMEEPHLKLVSIOMLVYRGMKVAVDGLALNPFYGS 926
 DB 25 GOTSASSVPPSAS-----TVAVDFASVTSSY--GKTIYVDELSEFVSAGS 67
 OY 927 QITSPFGNAGAGTTTMSILTLGTFPTSPGATVILGKDIKSEKSTIOMNGCPQHNVLDP 986
 DB 68 ECFGLGNGAGKSTIARMLGMCPPAGITVLGVVPARARLARARIGVPOQFNLDQ 127
 OY 987 MLTYEHIWEYARLKLSEKHVKAEMOMALDVLSPSSKLKSTLSQSGMOKISVALA 1046
 DB 128 EFTYRENILVGRYFGMSTQSEAVIPSL-LEFARLEKRNADAVSLSGGMKRLTMMARA 186
 OY 1047 FVGGSKVVIIDEPYAGDPSRKIGWE-LLKTYRGRTTLLSTHMDADVGLDPAITIS 1105
 DB 187 LINDPOLIVMDEPTTGIDPHARHLIWERLRLALRLANGKITIILTHFWEAEKLCDBLCVLE 246
 OY 1106 HGR 1108
 DB 247 KGR 249

RESULT 13

MODL_RHIGA STANDARD: PRT: 347 AA.
 AC P50332;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Modulation ATP-binding protein I.
 GN NOD1.
 OS Rhizobium galegae.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Rhizobium.
 OX NCBI_TaxID=399;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN=HAMBI 1174;
 RA MEDLINE=99403395; PubMed=10474187;
 RA Suominen L., Paulin L., Saano A., Saren A.M., Tas E., Lindstrom K.;
 RT "Identification of nodulation promoter (nod-box) regions of Rhizobium
 RT galegae.";
 RT FEBS Microbiol. Lett. 177:217-223(1999).
 CC -I- FUNCTION: FORMS, WITH NOD1, A MEMBRANE TRANSPORT COMPLEX INVOLVED
 CC IN THE NODULATION PROCESS. IT PROBABLY EXPORT A MODIFIED BETA-1,4-
 CC LINKED N-ACETYLGLUCOSAMINE OLIGOSACCHARIDE.
 CC -I- SUBCELLULAR LOCATION: Membrane-associated.
 CC -I- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: X87578; CAA60881.1;
 DR InterPro: IPR003593; AAA_ATPase.
 DR InterPro: IPR003439; ABC_Transport.
 DR Pfam: PF00005; ABC_tran; 1.
 DR ProDom: PD000006; ABC_transport; 1.
 DR SMART: SM00382; AAA; 1.
 DR TIGRfam: TIGR01288; AAA; 1.
 DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
 KM Modulation: Membrane; ATP-binding; Transport.
 FT NE_BIND 81 ATP (POTENTIAL).
 RT RT
 SO SEQUENCE 347 AA; 38435 MW; AC91210C44C9A6C CMC64;

Query Match 2.9%; Score 343.5; DB 1; Length 347;
 Best Local Similarity 30.0%; Pred. No. 3.8e-14;
 Matches 86; Conservative 58; Mismatches 126; Indels 17; Gaps 6;

OY 1873 REFIRPRV--NAKLSPLNDEDEDVARRRQRIILDDCGQNDILEIKELKTYRRRRKRAY 1929
 DB 7 REMLRPKTIAMDQASASARSNPREIKTGLEFPASASAPFMAIDQAVMYR--DKTV 64
 OY 1930 DRIVGTPPECCGLGVNAGAGSSTFKMLTGPTVTRGDAPLNNKLSLSNIEHWONG 1989
 DB 65 DLSFGVRAECGGLGGPNAGKSTIRMLLGWATPSACKISVLGIPVGKARLARASIG 124
 OY 1990 YCPODAITELTGRBHV---EFFALLRGVPEKGVKGEAATKLVKGEYAGNY 2045
 DB 125 VVSQFDLMEFVFRNLNLFVGYFOMSTRALIKLIPSLLEFA---OLEAKADRVSDL 180
 OY 2046 SGNKRRKLSFAMALIGPPVFLDEPTTGMDPKARFLMNCALSVKGRSVLTSHME 2105
 DB 181 SGNKRRKLTARALVNDPOLLIDEPPTGLDPPARHQIWERLRLIRGKITIILTTIMMD 240
 OY 2106 ECEALCTRMALVNGRFRCLG--SVQHLKRRFGDTIVRIAGSNPD 2151
 DB 241 EAERMCDRLCVLEGCRMIAEGPPLSLIDITG---CPVIEVYGGNPD 284

RESULT 14

MODL_HUMAN STANDARD: PRT: 1280 AA.
 AC P08183; Q12755; Q14812;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Multidrug resistance protein 1 (P-glycoprotein 1) (CD243 antigen).
 GN ABCB1 OR PGY1 OR MDR1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN 11
 RP SEQUENCE FROM N.A.
 RC MEDLINE=87028230; PubMed=2876781;
 RA Chen C.-J., Chin J.E., Ueda K., Clark D.P., Pastan I., Gottesman M.M.,
 RA Roninson I.B.;
 RT "Internal duplication and homology with bacterial transport proteins
 RT in the mdr1 (P-glycoprotein) gene from multidrug-resistant human
 RT cells.";
 RT Cell 47:381-389(1986).
 RL 12
 RN SEQUENCE FROM N.A.
 RP MEDLINE=90094448; PubMed=1967175;
 RA Chen C.-J., Clark D.P., Ueda K., Pastan I., Gottesman M.M.,
 RA Roninson I.B.;
 RT "Genomic organization of the human multidrug resistance (MDR1) gene
 RT and origin of P-glycoproteins.";
 RT J. Biol. Chem. 265:505-514(1990).
 RN 13
 RN SEQUENCE FROM N.A.
 RP MEDLINE=97190336; PubMed=9038218;
 RA Chen G., Duran G.E., Steger K.A., Lacayo N.J., Jaffrezou J.P.,
 RA Dumontet C., Sikic B.I.;
 RT "Multidrug-resistant human sarcoma cells with a mutant P-glycoprotein,
 RT altered phenotype, and resistance to cyclosporins.";
 RT J. Biol. Chem. 272:5974-5982(1997).
 RN 14
 RN SEQUENCE OF 1-234 FROM N.A.
 RA Smith A., Beck C., Gibson A.;
 RP Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.
 RN 15
 RP SEQUENCE OF 178-215 AND 800-856 FROM N.A.
 RX MEDLINE=90290529; PubMed=1972623;
 RA Gekeler V., Weger S., Probst H.;
 RT "mdr1/P-glycoprotein gene segments analyzed from various human
 RT leukemic cell lines exhibiting different multidrug resistance


```

QY 1212 AAKGAFVLEFHEIDRLSDIGISSYGISETTLEIFLKVAESGDAENSDGTLFARRN 1271
Db 666 -----
QY 1272 RAQFDKQSCRLPFTEDDAADPNDSIDIPESRETDLGMDGSGYQWKMLTQOQVFA 1331
Db 669 RRSVGSQAQDRKLTSTKEAL--DESLPP-----VSWRMLKMLNLE 707
QY 1332 LMKRLILARRSRKGFQAOYLPAVFCIALVSLVPPPGKYPSLELOFMVNEOYFV 1391
Db 708 --WPEFV-----GVFAITINGILOPAFAIIFSKII----- 736
QY 1392 SNADEDTGTELLNALTKDPGEGTROMEGNPIDPTPCQAGEPEWMTAPVOTIMLPON 1451
Db 737 -----GVFTR----- 741
QY 1452 GNMWIMQNPSPACQSSSKIKMLPVCPCGAGLPPPOKQONTADLLODLGHRNSDYLVK 1511
Db 742 -----IDPEFTKRONSLFSL-----FLAL 762
QY 1512 TYVOILAKSLKNIWNEFRYGSFSLGVSNVQALPPSOQVNDALIKOMKHILKLAQSS-A 1570
Db 763 GLISFI-----FLOGFTFGKA-----GETILTKLRMYWFRSMLRQDVSWF 804
QY 1571 DRFLNSLGRMTGLDRNNKVFENKGMHAISFLVNNALILRANLQKGENPSHYGT 1630
Db 805 DDEKNTGALITRL-ANDAAOV---KG--AIGSRLAVITQNI-ANLGTI-----II 849
QY 1631 AFNHPPLNLTQOQSEVALMTSDVLSVCIFFAMSPFASVVLIOFRYSKAKHLOFT 1690
Db 850 SFYGMQLT-----LLILAVPLIATAGVEMK-----MLSQALKDKKELE-- 891
QY 1691 SGKPIYIWNLSNFWDMQNVVPAVLVITIFCPOOK---SVVSSSTNPLVALL--LLY 1745
Db 892 GACKIATEAIEN-----RIYVSLTQOKFEHMYAQSLOVYPNSLRKAHIF 938
QY 1746 G--WSTPLMYPASFEYKIPSTAYVVLVSVNLFIGINSVATVLELFTDNKLNINDIL 1803
Db 939 GIFPSTQAMMFYSAGCGRFAYIVA-----HKLMSFEDVL 975
QY 1804 KSVFLFPFPCIGRCILDMVKNQAMADALERGENRFV---SPLSMDLVGRNLFAMAV 1858
Db 976 ---LVESAIVGAMAVQGVSSFADYAKAKISAHIIMIIEKPLLDISTEGLMNTL 1931
QY 1859 EGVEFELLVLIOYRFRPR-PAVAKSPLDEDEVREROKILDGCGONDILTEIKEL 1917
Db 1032 EGNVTFGEV---ENVPTRPDIPLYGLS-----LEVKK- 1062
QY 1918 TTIYRRKKPAVDRIQVIGIPGECFGLGVNAGKSSIFKMLTGTDTVTGDAFLNNSI 1977
Db 1063 -----GOTLALVSSGCGKSTVYVOLLERFYPLAGKYLDOKEI 1101
QY 1978 LS-NIHEVHOMNGYCPQ---FD-----AITELLIGREHVEFFALLRGV 2016
Db 1102 KRLNQMILRAHGLIYSDPILFDCSIAENIAYGDSNRVSOEELIVRAKANIHAFTESL 1161
QY 2017 PEKEVKGVEMAIRKGLVYKGEKYAGNYSQGNKRKLSTAMALIGCPVYVFLDEPTGMD 2076
Db 1162 PNKYSTKVGDKGTO-----LSGQKQRIAIRALVROPHILLDEATSAID 1207
QY 2077 PKARRELINCLLSVKEGSRVVLTSHECEALCTRMALVNGRRPLGSGVOHLKNRG 2136
Db 1208 TESKRVQOE-ALDKARREGRCIYIAHRLSTION-ADLIVFONGRVKENGHTQOQLAOKG 1265
QY 2137 DGYTIVIRIAGS 2148
Db 1266 IYFSMVSVOAGT 1277

```

```

RESULT 15
NODI_RHIME
ID NODI_RHIME
AC 052618;
STANDARD;
PRT: 355 AA.

```

```

DT 30-MAY-2000 (Rel. 39, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Nucleotide sequence of nodJ region of Rhizobium meliloti pSymA.
GN NODI OR RA0472 OR SMA0864.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OG Plasmid pSymA (megaplasmid 1).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
CX NCBI_TaxID=382;
RN 111
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396509; PubMed=11481432;
RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
RA Barley-Hubler F., Bowser L., Capela D., Galibert F., Gouzy J.,
RA Gurjal M., Hong A., Huzar L., Hyman R.W., Kahn D., Kahn M.L.,
RA Kaiman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,
RA Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;
RT "Nucleotide sequence and predicted functions of the entire
RT Sinorhizobium meliloti pSymA megaplasmid".
RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).
RN 121
RP SEQUENCE OF 143-355 FROM N.A.
RC STRAIN=1021;
RA Barnett M.J., Long S.R.;
RT "Nucleotide sequence of nodJ region of Rhizobium meliloti pSymA.
RT Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: FORMS, WITH NODJ, A MEMBRANE TRANSPORT COMPLEX INVOLVED
CC IN THE NODULATION PROCESS. IT PROBABLY EXPORT A MODIFIED BETA-1,4-
CC LINKED N-ACETYLGUCOSAMINE OLIGOSACCHARIDE.
CC -1- SUBCELLULAR LOCATION: Membrane-associated.
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE007237; AA065130.1; -
CC EMBL: AF043118; AB97762.1; -
CC InterPro: IPR003593; AAA_ATPase.
CC InterPro: IPR003439; ABC_transportr.
CC Pfam: PF00005; ABC_tran; 1.
CC ProDom: PD000006; ABC_transportr; 1.
CC SMART: SM00382; AAA; 1.
CC TIGRFAMs: TIGR01288; nodJ; 1.
CC PROSITE: PS00211; ABC_TRANSPORTER; 1.
CC Nucleotide sequence. Membrane; ATP-binding; Transport; Plasmid;
KW Complete proteome.
FT NP_BIND 89
FT SEQUENCE 355 AA; 39268 MW; 4DC869D98C35DC CRC64;

```

```

Query Match 2.8%; Score 331.5; DB 1; Length 355;
Best Local Similarity 33.9%; Pct. No. 2.3e-13;
Matches 74; Conservative 52; Mismatches 85; Indels 7; Gaps 4;

```

```

QY 891 EPHLKIGVSIQNLVKKYRDGMKVAADGALNFEQOITPSFGHNGAGTMTSTIGLF 950
Db 52 KPT---VAIDVASVTKSYGD--KPVINGLSFIVVAGCEGLGPGACKSTTRHILGMT 106
QY 951 PPTSGTAYIIGKDIRSEMSITRNIGVCGQHNVLDMLTVEHIFVYRLKLGSEKHVA 1010
Db 107 TPGGGETIVYGVPPSRKRLARRIGVWPQFNDLEFTVRENILVFRYRPMSTREIEA 166
QY 1011 EMDQALDVGLPSSKLSKTSQSLSGGMQRKLSVALAFVGSKVVLIDPPTGVPYSGRG 1070
Db 167 VIPSL-LEFARLENKADARVSDLSGGMKRRRLIARALINDPOLLDLDEPTTGDDPHARHL 225
QY 1071 IWEILLK-YKQGCTIILSIHMDADVLDGRIATISHG 1107

```

Fri Mar 7 09:16:54 2003

us-09-595-526c-2.rsp

Page 21

Db 226 IWERLRSILARGTILLTHIMEAERLCODRLCVLEAG 263

Search completed: March 7, 2003, 08:48:05
Job time : 61 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 7, 2003, 08:44:14 ; Search time 64 Seconds
(without alignments)
3396.249 Million cell updates/sec

Title: US-09-595-526C-2
11797

Sequence: 1 MACWPOLRLMLKMLTFRRR.....VDVAVLTSFLQDEKVESTV 2261

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 segs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: PIR_73:*

1: PIR1: *
2: PIR2: *
3: PIR3: *
4: PIR4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	10922	92.6	2201	2 A54774	ATP binding cassette
2	3345.5	28.4	1529	2 A59189	ATP-binding cassette
3	3129.5	26.5	1472	2 B54774	ATP binding cassette
4	2645.5	22.4	1704	2 S71363	probable ATP-bind
5	2642.5	22.4	1704	2 A59188	ATP-binding cassette
6	2060	17.5	1802	2 T33783	hypothetical prote
7	1984	16.8	1816	2 A84845	probable ABC trans
8	1863.5	15.8	1447	2 T15200	hypothetical prote
9	1792	15.2	1317	2 C88925	protein F33E11.4 [
10	1542	13.1	1758	2 F88559	ATP-binding cassel
11	1538.5	13.0	1704	2 T42749	transport protein
12	1538.5	13.0	1767	2 S60124	hypothetical prote
13	1402	11.9	1246	2 T00826	hypothetical prote
14	1158.5	9.8	1564	2 T27121	hypothetical prote
15	1013.5	8.6	373	2 T42150	hypothetical prote
16	1010.5	8.6	1431	2 T42467	hypothetical prote
17	846	7.2	269	2 T46457	hypothetical prote
18	788.5	6.7	1011	2 T07712	probable ABC-type
19	693	5.9	900	2 T07717	probable ABC-type
20	667	5.7	895	2 T07714	probable ABC-type
21	656.5	5.6	722	2 T07716	probable ABC-type
22	645	5.5	925	2 T07713	probable ABC-type
23	624.5	5.3	1336	2 T18288	ATP transport prot
24	485	4.1	196	2 T12512	hypothetical prote
25	432	3.7	339	2 S74048	probable daonorubi
26	428	3.6	664	2 T07715	probable ABC-type
27	426.5	3.6	324	2 C71081	probable resistanc
28	421	3.6	328	2 E75108	daonorubidin resis
29	411	3.5	330	2 S27707	daonorubidin resis

30	410.5	3.5	310	2 C96929	ABC-type multicludg
31	410.5	3.5	327	2 D72257	hypothetical prote
32	410	3.5	314	2 D97318	ABC-type MDR trans
33	409.5	3.5	398	2 C69485	daonorubidin resis
34	403.5	3.4	297	2 A81816	ABC transporter (A
35	396.5	3.4	333	2 D72492	probable ABC trans
36	395.5	3.4	246	2 S75436	hypothetical prote
37	390	3.3	301	2 E72384	ABC transporter, A
38	388.5	3.3	310	2 E69920	ABC transporter (A
39	388	3.3	311	2 G69803	probable ABC-type
40	386.5	3.3	727	2 T07718	ABC-type MDR trans
41	386	3.3	259	2 H97301	probable dtra prot
42	386	3.3	331	2 D70984	ABC transporter (A
43	385	3.3	350	2 B69065	ABC transporter (A
44	384	3.3	331	2 S73019	daonorubidin resis
45	380	3.2	312	2 H97342	ABC-type MDR trans

ALIGNMENTS

RESULT 1

A54774
ATP binding cassette transporter ABC1 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 05-Apr-1995 #sequence_revision 05-Apr-1995 #text_change 02-Feb-2001
C:Accession: A54774
R:Luciani, M.F.; Denizot, F.; Savary, S.; Mattei, M.G.; Chimini, G.
Genomics 21, 150-159, 1994
A:Title: Cloning of two novel ABC transporters mapping on human chromosome 9.
A:Reference number: A54774; M01D:94375008; PMID:8088782
A:Accession: A54774
A:Molecule type: mRNA
A:Residues: 1-2201 <LUC>
A:Cross-references: GB:X75926; NID:9495256; PID:CA53530.1; PID:9495257
C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology
C:Keywords: ATP; duplication; nucleotide binding; P-loop
F:856-1047/Domain: ATP-binding cassette homology <ABC1>
F:873-880/Region: nucleotide-binding motif A (P-loop)
F:1869-2060/Domain: ATP-binding cassette homology <ABC2>
F:1886-1893/Region: nucleotide-binding motif A (P-loop)

Query Match	Score	92.6%	10922	DB 2	Length	2201
Best Local Similarity	94.9%	Pred. No. 0				
Matches 2089	Conservative 55	Mismatches 57	Indels 0	Gaps 0		
OY	61	MPAGTLPMWOGIICNANPCFRYPGAPGVGNFNKSIYARLFSARILLISQKDT	120			
DB	1	MPAGTLPMWOGIICNANPCFRYPGAPGVGNFNKSIYARLFSARILLISQKDT	60			
OY	121	SMKDKRYLRTLDQIKKSSNKLQDLVUNEFPSGLYHNLSLPKSTYDKMLRADVTLH	180			
DB	61	SIKDKHKLRLMIRQIKPNSNMLQDLVUNEFPSGLYHNLSLPKSTYDKMLRADVTLH	120			
OY	181	KVFLGQYGLHLTSLCSKSGSEMTOLDQEVSELCGPRKELAAERVLRYNMDILKPTL	240			
DB	121	KVFLGQYGLHLTSLCSKSGSEMTOLDQEVSELCGPRKELAAERVLRYNMDILKPTL	180			
OY	241	RTLNSFSPFSKELAEAVTKLLSLGLTADLEFSMSWSMROEWFLLTVNNSSSSTOI	300			
DB	181	TKLNTSHLPQHLAEATVTLDSLGLADELFTKSMDSMRQEWMLTVNNSSSSTOI	240			
OY	301	YOAVSRIVCGHREGGGLIKSLNMYEDNNYKALFGNGTEDEAFPDNNTTYCDMLK	360			
DB	241	YOAVSRIVCGHREGGGLIKSLNMYEDNNYKALFGNGTEDEAFPDNNTTYCDMLK	300			
OY	361	NESSPLSLRIWALKPLLVGKILYTPDTPATROYMAEVNKTFOELAVFDLQGMVELS	420			
DB	301	NESSPLSLRIWALKPLLVGKILYTPDTPATROYMAEVNKTFOELAVFDLQGMVELS	360			
OY	421	PKIWTFMENSQMDLVKRLMDSRDNDHFWEQDGLDMTADDIYAFLAKKPEDEVQSSNGS	480			
DB	361	PKIWTFMENSQMDLVKRLMDSRDNDHFWEQDGLDMTADDIYAFLAKKPEDEVQSSNGS	420			

QY 481 VYTRAPFNETNATPTISRFMECVNLKLEPATEVWLINXSNELLDERKFMAGIYFTG 540
 |||||
 Db 421 VYTRAPFNETNATPTISRFMECVNLKLEPATEVWLINXSNELLDERKFMAGIYFTG 480
 |||||
 QY 541 ITPGSIELPHHKKYKIRMDIDVNEFKTKIKGYPMPGRADPEMDKRYWNGEAYLDDV 600
 |||||
 Db 481 ITPDSVLEPHHKKYKIRMDIDVNEFKTKIKGYPMPGRADPEMDKRYWNGEAYLDDV 540
 |||||
 QY 601 EQAITRVLTGFKTKGYVMQOMPFCYVYDIFLRVMSKMDLFMTLAMIYSAVAIIKGIY 660
 |||||
 Db 541 EQAITRVLTGFKTKGYVMQOMPFCYVYDIFLRVMSKMDLFMTLAMIYSAVAIIKGIY 600
 |||||
 QY 661 YEKERARKETMRIMGLDMSILMFESWISLIPILVYAGLLVILIKGLNLPSPDPVVEY 720
 |||||
 Db 601 YEKERARKETMRIMGLDMSILMFESWISLIPILVYAGLLVILIKGLNLPSPDPVVEY 660
 |||||
 QY 721 FLSVAVVYIIQCLIFSLFSTRANLAAACGGIYFTLYLPVLCVAMQDYVGFTLKIFAS 780
 |||||
 Db 661 FLSVAVVYIIQCLIFSLFSTRANLAAACGGIYFTLYLPVLCVAMQDYVGFTLKIFAS 720
 |||||
 QY 781 LLSVPAFGCEYFALFEEOGIGVOMNLPESVYEDGFMULTISIMLEPDTFLYGMVW 840
 |||||
 Db 721 LLSVPAFGCEYFALFEEOGIGVOMNLPESVYEDGFMULTISIMLEPDTFLYGMVW 780
 |||||
 QY 841 YIEAVFPGQYGIIPRMYFPCTKSYWFGESDEKSHPGSNOKRMSSEICMEDEPTHLKIGVS 900
 |||||
 Db 781 YIEAVFPGQYGIIPRMYFPCTKSYWFGESDEKSHPGSNOKRMSSEICMEDEPTHLKIGVS 840
 |||||
 QY 901 IONLVYRDMGKVAVDGLANLFEQGITFSLHNGAGKTTNLTGLTGPPTSGTAYIL 960
 |||||
 Db 841 IONLVYRDMGKVAVDGLANLFEQGITFSLHNGAGKTTNLTGLTGPPTSGTAYIL 900
 |||||
 QY 961 GKDIRSEKSTIRONLGVCPDHNVLFDMLTVEEHIMFARLKGISEKHVAEMEQNALDVG 1020
 |||||
 Db 901 GKDIRSEKSTIRONLGVCPDHNVLFDMLTVEEHIMFARLKGISEKHVAEMEQNALDVG 960
 |||||
 QY 1021 LPSKLSKTSQSLSGQKRLSVALAFVGGSKVYILDEPTAGVDSRRKGIWELLTKYRQ 1080
 |||||
 Db 961 LPSKLSKTSQSLSGQKRLSVALAFVGGSKVYILDEPTAGVDSRRKGIWELLTKYRQ 1020
 |||||
 QY 1081 GRTIILSTHMDADVGDGDIATISHGKLCYSSSLFLKNOIGTYILTVLKKDVESLS 1140
 |||||
 Db 1021 GRTIILSTHMDADVGDGDIATISHGKLCYSSSLFLKNOIGTYILTVLKKDVESLS 1080
 |||||
 QY 1141 SCRNSSTVYTLKEDSVSSSDAGLSGDHESDITLIDVSAISMLIRKHVSEARLVEDI 1200
 |||||
 Db 1081 SCRNSSTVYTLKEDSVSSSDAGLSGDHESDITLIDVSAISMLIRKHVSEARLVEDI 1140
 |||||
 QY 1201 GHEITVYLPYEAKEGAFVELFHEIDRLSDIGISSYGISETTLEETFLKAESGVDAE 1260
 |||||
 Db 1141 GHEITVYLPYEAKEGAFVELFHEIDRLSDIGISSYGISETTLEETFLKAESGVDAE 1200
 |||||
 QY 1261 TSGCITLPPARNRRAFCDKOSCLRPTEDDAADPNDSIDIPESRETDLSGMDGKGSYQVK 1320
 |||||
 Db 1201 TSGCITLPPARNRRAFCDKOSCLRPTEDDAADPNDSIDIPESRETDLSGMDGKGSYQVK 1260
 |||||
 QY 1321 GMLTQOQVALLMKRLILARRSRKGFAQIYVLPVAFVCIALVSLVPPGKPSLELQ 1380
 |||||
 Db 1261 GMLTQOQVALLMKRLILARRSRKGFAQIYVLPVAFVCIALVSLVPPGKPSLELQ 1320
 |||||
 QY 1381 PMMYNEQYTVFSNDAPEDTGTELELNAITKDPGFTRCMEBNPJPDPPCOQGEEMWTAP 1440
 |||||
 Db 1321 PMMYNEQYTVFSNDAPEDTGTELELNAITKDPGFTRCMEBNPJPDPPCOQGEEMWTAP 1380
 |||||
 QY 1441 VQOTIMDLFONGNMTMNPSPACQSSSKIKKMLPVCPGAGGLPPQKQNTADILQDL 1500
 |||||
 Db 1381 VQOTIMDLFONGNMTMNPSPACQSSSKIKKMLPVCPGAGGLPPQKQNTADILQDL 1440
 |||||
 QY 1501 TGRNISDYLVTYVQIILAKSLKNKIMVNEFFYGGFISGVSTQALPPSQOEVNDAIKOMKK 1560
 |||||
 Db 1441 TGRNISDYLVTYVQIILAKSLKNKIMVNEFFYGGFISGVSTQALPPSQOEVNDAIKOMKK 1500
 |||||

QY 1561 HILKADSSADBEFLNSLGRFMTGLDTRNNKVVWFNNKGHAISSFLVNNAILRANLQK 1620
 |||||
 Db 1501 LKLTKIDTSADBEFLNSLGRFMTGLDTRNNKVVWFNNKGHAISSFLVNNAILRANLQK 1560
 |||||
 QY 1621 GENPSHYGITAENHNLJTKOOLSEVALMTTSYVULVASICIVAFMSVPSVFFLIQER 1680
 |||||
 Db 1561 GENPSHYGITAENHNLJTKOOLSEVALMTTSYVULVASICIVAFMSVPSVFFLIQER 1620
 |||||
 QY 1681 VSKAHILQIFISGVKRVYIWSLNFYWDKCNVVPATVLIIFICFOOKSVYSSTNLPVAL 1740
 |||||
 Db 1621 VSKAHILQIFISGVKRVYIWSLNFYWDKCNVVPATVLIIFICFOOKSVYSSTNLPVAL 1680
 |||||
 QY 1741 LLLYGMSTPLPAMPASVFEKIPSTAYVYLTSYNLFGINGSAATVLELFDNKNLNTN 1800
 |||||
 Db 1681 LLLYGMSTPLPAMPASVFEKIPSTAYVYLTSYNLFGINGSAATVLELFDNKNLNTN 1740
 |||||
 QY 1801 DILKSVFLIPFHCLGRLIDMVKKNQAMADALERFGENRVSPLSMDLGRNLFMAVVG 1860
 |||||
 Db 1741 DILKSVFLIPFHCLGRLIDMVKKNQAMADALERFGENRVSPLSMDLGRNLFMAVVG 1800
 |||||
 QY 1861 VVFFLITVLIQYREFIRPRVNAKLSPINDEDEDVAREQRILIDGGONDLEIKELTKI 1920
 |||||
 Db 1801 VVFFLITVLIQYREFIRPRVNAKLSPINDEDEDVAREQRILIDGGONDLEIKELTKI 1860
 |||||
 QY 1921 YRRKRKPAVDRLCVGLIPGCEFGILGVNAGKSTFKMLTGTVTYRGDAFLNKSIIISN 1980
 |||||
 Db 1861 YRRKRKPAVDRLCVGLIPGCEFGILGVNAGKSTFKMLTGTVTYRGDAFLNKSIIISN 1920
 |||||
 QY 1981 IHEYHOMNGYCPDFDITELLTGREHVEFFALLRGVPEKVGKVBEMATRKILGVKYEK 2040
 |||||
 Db 1921 IHEYHOMNGYCPDFDITELLTGREHVEFFALLRGVPEKVGKVBEMATRKILGVKYEK 1980
 |||||
 QY 2041 YAGNYSQGNKRLSTAMALIGCPVVEITDEPTTGMDPKARRFLMCAISVKEGSSVLT 2100
 |||||
 Db 1981 YAGNYSQGNKRLSTAMALIGCPVVEITDEPTTGMDPKARRFLMCAISVKEGSSVLT 2040
 |||||
 QY 2101 SHMECEALCTMAIMVNGRFRCLGSVOHLKRRFGDGYTIVYRAGSNPLDKVPDFFG 2160
 |||||
 Db 2041 SHMECEALCTMAIMVNGRFRCLGSVOHLKRRFGDGYTIVYRAGSNPLDKVPDFFG 2100
 |||||
 QY 2161 IAPPGSVLKEKRRNMLQYOLPSSLSIARIFSIISOSKRLIHIEVYSQTTLDQVYVNF 2220
 |||||
 Db 2101 IAPPGSVLKEKRRNMLQYOLPSSLSIARIFSIISOSKRLIHIEVYSQTTLDQVYVNF 2160
 |||||
 QY 2221 AKQOSDDHLKDLJHKNOTVVDVAVLTSFLQDEKVKESYV 2261
 |||||
 Db 2161 AKQOSDDHLKDLJHKNOTVVDVAVLTSFLQDEKVKESYV 2201
 |||||

RESULT 2

A59189
 Atp-binding cassette transporter - human (fragment)
 N:Alternate names: KIA1062 protein
 C:Species: Homo sapiens (man)
 C:Date: 18-Feb-2000 #sequence revision 18-Feb-2000 #text_change 02-Jun-2000
 C:Accession: A59189
 R:Kikuno, R.; Nagase, T.; Ishikawa, K.; Hirotsawa, M.; Miyajima, N.; Tanaka, A.; Kotan
 DNA Res. 6, 197-205, 1999
 A:Title: Prediction of the coding sequences of unidentified human genes. XIV. The com
 A:Reference number: 222961; MUID:9397452; PMID:10470851
 A:Accession: A59189
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-1529 <KIK>
 A:Cross-references: GB:AB028985; MID:q5689460; PIDN:BAAB3014.1; PID:q1046841; PID:q56
 A:Experimental source: chromosome 9; clone hJ03579; clone 11b pBluescriptII SK plus;
 A:Genetics:
 A:Map position: 9
 A:Note: KIA1062
 C:Superfamily: unassigned Atp-binding cassette proteins; Atp-binding cassette homolog
 Query Match 28.4%; Score 3345.5; DB 2; Length 1529;
 Best Local Similarity 46.7%; Pred. No. 4.1e-208;

Matches 721; Conservative 207; Mismatches 381; Indels 235; Gaps 33;

```

QY 836 GWTWYIENAFVGGYGRPMWYFPCSTKSYMFE---ESDEKSHFGSNOKRMS-----ETC 887
   1:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 GLTWTYIENAFVGGYGRPMWYFPCSTKSYMFE---ESDEKSHFGSNOKRMS-----ETC 60
QY 888 -----MEEPHKLKGVSIQNLVYVYRDGMVAVDGLALNFEGQITSPFGHN 935
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 AMESRREETROMEEPHLPLVVCVDKLTYYKDKDKLALNKLSINLEYQVVSFLGHN 120
QY 936 GACKTTMTSLILGPPPTSGATYILGKDIRSEMTIRONLGVCPQHNVLPLMLVEEHIM 995
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 GACKTTMTSLILGPPPTSGATYILGKDIRSEMTIRONLGVCPQHNVLPLMLVEEHIM 180
QY 996 FYARLGLSEKHVAKMEOMALDVGIPSSKLKSKTSQSLSGCMORLKVLAFAVGGSRAL 1055
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 FYARLGLSEKHVAKMEOMALDVGIPSSKLKSKTSQSLSGCMORLKVLAFAVGGSRAL 239
QY 1056 LDEPTAGVDPYSRGIMWELLKRYRGRTIILSTHMDADVLGDRITAIISHGKLCVSS 1115
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 240 LDEPTAGVDPYSRGIMWELLKRYRGRTIILSTHMDADVLGDRITAIISHGKLCVSS 299
QY 1116 LELKNOLGTGYTLVKKDVESLSCRNSSSTVSLKEDVSQSSDAGLSDHESDT 1175
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 300 LELKNOLGTGYTLVKKDVESLSCRNSSSTVSLKEDVSQSSDAGLSDHESDT 1175
QY 1176 LTIDVS--ATSNLIRKHVSEARLVEDIGHETVLPYEAKEGAFVLEHEIDRLSDLG 1233
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 338 PLSSCELSQVQFTRKHVAGCLLVSDTSTLSYLPSSEAKKAGFRLRQHLERSIDALH 397
QY 1234 ISSYGISETTELEIFLYAAE-----SGYDA-ETSDGTLP----- 1267
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 398 LSSFGIMDTLEEVFLKYSSEDOLENSSEADYKESRDVLPAGBAGPSGCHAGNARCS 457
QY 1268 -----ARRRRA-FGDKOSCLRFTEDDADPNDSIDIPESRETDLS 1309
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 458 ELTQSOASLOSASSVSKRGDEGAGTYDYGRLP-DNODPD--NSLOVEAEALS 514
QY 1310 GMDGSGYOVKGMKLTQOQFVALLMKRLILARRSKRGFAQVLPVAVFCIALVFSILP 1369
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 515 RV-GGCSRKLDGMLKVFQFHLILKRFHCARNSKALFSLDLPFAFVCAAMVALSV 573
QY 1370 PRGKPSLELOPWNMYNEQT-----FVSNDAPE-----DTGTELLNALTKDPGF 1414
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 574 EIGDLPPLVLSPOYH-NITOPRGNFYXANDEREYLRISPDASPOGLVSTRLPSGV 632
QY 1415 GTRCM-----EGNPT----- 1424
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 633 GATCVLAKSPANGSLGPTLNLSSGESRLAARFDSMCLESFTQGLPLSNFYVPPSPAPS 692
QY 1425 -----PD-----TROAGEEETTPAR-VROTIMDLFONGNTMQNPSPACQSSDK 1469
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 693 DSPASPEDLOAMNVSILPPTAGEMTTSAPSLRVLREVVR-----CICSAQ 740
QY 1470 IKKMLRVCPRGAGLPPRPKONTADTLODLTGRTISDYLKVTQYQITAKSLKNTWNE 1529
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 741 TGGS---CPSSVGG-HPQGRVVTGDIILMDITGHNVSEVLEFTSDRF-----RL 785
QY 1530 FRIGGESLGSNTQALPRPSEVNDATIKOMKHKLKLAKSSADREFLNSLGRFTGLDTRNN 1589
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 786 HRYGATIFG---NVLKSTIPASFTGRAPRWYK-----IAVRR 820
QY 1590 VKVFNKNGHAISSFLNVLNNAILRANLQKGE-NPSHYGTAENHPLNLTKQOLSEVAL 1648
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 821 AQVYNNKGYHSMPTYLNSLNNALIRANLPRSKGNPAAGYITVNPMMKTSASLS-LDY 879
QY 1649 MTSVDVLVLCVTFPAMSPFAPSVVFLIOBRYSKAKHLOFISGVAPVLYLMSFENWDMC 1708
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 880 LLOGDVLVIAFTIVAMSFVAPSVVFLVLEKSTKAKHLOFVSGCNPIIYMLANVWML 939
QY 1709 NYVPATVITITICFOOKSVSSSTNLPLVALLLXGMSITPLMPASVETFTSTYAT 1768
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 940 NYLVATCCVILLVFDLPATYSTNPAVALSLFLYGWSITPTIMVPASWFEVPSAVY 999

```

```

QY 1769 VFTSVNLTGINGSVATVLELFT-DNKLNNDILKSVFLFEPHFLGKLDIWNKQA 1827
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1000 FLVLNLELTGATVATVLELFDLEHNDLKVNSYLSKSLFLPNNYLGHLMEMAYNE 1059
QY 1828 MADLALERGE-NREYSPLSMDLVGRNLFAMAVEGVFPLITVILQYFFLRPPRYNAKLS 1886
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1060 INEYAKIQOPKMSSPPEMDLVTRGILVAMAVEGVGLLITIMQVNFRLRPPRMVSTK 1119
QY 1887 PLNDEEDVRRERORILDOGGONDLLEIKELTKLYRRK---RKPAVDRIQVIRPGEFCF 1943
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1120 PVED-DVDVAASERORVLRDADNDKVKLEMLTKYKSKIRIILAVDRICGVARGCFG 1178
QY 1944 ILGVNAGKSPFKMLTGTPTVTRGDAFLNKNSILSNIEHVNQMGYCPQFDATIELTG 2003
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1179 LUGVNGAGKTSFTFKMLTGDESTTGGEAFVNGHSVLKELLQVOSLSGYCPOCDALDELTA 1238
QY 2004 REHVEFFALLRGVPEKEVGKGBEMAIRKGLVYGEKAGANSGCNKRKLSMAALGSP 2063
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1239 REHLOLYTRLRGISWMDARVYKMALEKLETKYADKPAGTYSGGNKRKLSALALIOY 1298
QY 2064 PVVFLDEPTGMDPKARRFLMNCALSVKREGRSVLTSHECEBALCTRMALVNGRFR 2123
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1299 AEFPLDEPTGMDPKARRFLMNCALSVKREGRSVLTSHECEBALCTRMALVNGRFR 1358
QY 2124 CLGVSQHLKRNFGDGYTIVRIAGSNPDLKPVODFFGLAFPPGSYLKEKHRNMLQVLPSS 2183
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1359 CLGSIQHLKRNFGDGYTIVRIAGSNPDLKPVODFFGLAFPPGSYLKEKHRNMLQVLPSS 1417
QY 2184 LSSLARIFSLISOSKRLNHEEDYSQTTLDQVFNFAVDQSD 2227
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1418 HISLAOVFSKMEQVSGVLGIEDISVSTTIDNVFNVPARKQSDN 1461

```

RESULT 3

B54774
 ATP binding cassette transporter ABC2 - mouse (fragment)
 C:Species: Mus musculus (house mouse)
 C:Date: 23-Mar-1995 #sequence_revision 05-Apr-1995 #text_change 02-Feb-2001
 C:Accession: B54774
 R:Luciani, M.F.; Denizot, F.; Savary, S.; Mattel, M.G.; Chmialni, G.
 Genomics 21, 150-159, 1994
 A:Title: Cloning of two novel ABC transporters mapping on human chromosome 9.
 A:Reference number: A54774; MUID:94375008; PMID:808782
 A:Molecule type: mRNA
 A:Accession: B54774
 A:Residues: 1-1472 <LUC>
 A:Cross-references: GB:X75927; NID:q495258; PIDN:CA53531.1; PID:q495259
 C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homolog
 C:Keywords: ATP; nucleotide binding; P-loop
 F:44-234/Domain: ATP-binding cassette homology <ABC1>
 F:51-68/Region: nucleotide-binding motif A (P-loop)
 F:108-1300/Domain: ATP-binding cassette homology <ABC2>
 F:1126-1133/Region: nucleotide-binding motif A (P-loop)

Query Match 26.5%; Score 3129.5; DB 2; Length 1472;
 Best Local Similarity 46.0%; Pred. No. 4.1e-194;
 Matches 676; Conservative 207; Mismatches 377; Indels 211; Gaps 28;

```

QY 888 MEEPHKLKGVSIQNLVYVYRDGMVAVDGLALNFEGQITSPFGHNAGKTTMTSLT 947
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 16 MEEPHLPLVVCVDKLTYYKDKDKLALNKLSINLEYQVVSFLGHNAGKTTMTSLT 75
QY 948 GLPPTSGATYILGKDIRSEMTIRONLGVCPQHNVLPLMLVEEHIMFYARLGLSKH 1007
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 76 GLPPTSGATYILGKDIRSEMTIRONLGVCPQHNVLPLMLVEEHIMFYARLGLSKH 135
QY 1008 VKAEMOALDVGIPSSKLKSKTSQSLSGCMORLKVLAFAVGGSRALILDEPTAGVDPY 1067
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 136 IRKEDKMIEDLEL-SNKRHSLVQTLGSKMKRLSAIAFAVGGSRALILDEPTAGVDPY 194
QY 1068 RRGIMWELLKRYRGRTIILSTHMDADVLGDRITAIISHGKLCVSSFLKNOLGTGY 1127
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

Db 195 RRAIMDLILKYKGRFTIILSTHMDADLLDRIALIIISHGKLCCKGSPLELKGAYXDXGR 254
Qy 1128 LTLVKKDVSSISCRSSSTVSVYLKKEDSVQSSDAGLGDHSDTLTLDVSAISNLT 1187
Db 255 LTLVQPAPEPTSOBPGLASSPSCPRILSSCSFEO-----VSQFI 294
Qy 1188 RKNVSEARLVEDIGHELTYVLPYEAKEGSAFVELFHEIDRLDLSGISYGETTLEET 1247
Db 295 RKNVASSLTVSDTSTELSYILPSPAVKGAERLFOOLEHSLDALHLSFGIMDTLIEV 354
Qy 1248 FLKAAEF-----SGVA-ETSDGTLPARNRRAFGDKOSCLPFTF----- 1287
Db 355 FLKVSSEDSLENSADYKESRKVDLPGAEGLTAVGGQGNLARGSELAAQSOASISQASS 414
Qy 1288 -----DDADPNDSDIDPESRTDLSGMDKGSYQYKMWL 1324
Db 415 VGSARGEETGYSDGYDYPRLFDNLQDDP--NVSLOEAEMLAAY--GQGSRLKLGWML 471
Qy 1325 TQOQFVALLMKRLLIARSKKGFQAIYLPVAVFCIALFSLVPPGKYPSLELQPMX 1384
Db 472 KMRQFHLVKKRHCARSKALCSOILPAPFVCAVMVALSVPEIGDLPPLVLSPOY 531
Qy 1385 NEOYT-----FVSNDAPE-----DTGTELLNALTQDPGCTRCM----- 1419
Db 532 H-NVTPQRCNFTYANEFQOEYRLSPASPQQLVSTFRLPSPGVGATCVLKSPANGSLG 590
Qy 1420 -----EGNPI-----PD----- 1426
Db 591 PMINISGESRLLAARFPDSMCLSEFTQGLPLSNFVPPPPSPASDSVYXDEDSLOAWN 650
Qy 1427 --TPCQAGEEWTTAP--VPQTINDLFQNGNMTMNPSPACOSSSDKIKKMLVCPGAG 1483
Db 651 MSLPPTAGETWTSAPSLRVLVHEVR-----CTCSAQGGTGS--CPSSVGG 695
Qy 1484 LPPQKQNTADLIDLTGCRNISDYLYVVOILAKSLKNKIVNEPRYGGESLGVSNTQ 1543
Db 696 -HPPQRYVYTGILDTIGHNVEYLFTSDRF-----RLRHYGATITG--NVQ 741
Qy 1544 ALRPSQEVDAIKQKKHLAKDSADRFLNISLGRFTGIDTRNNVYVFNKMGHATS 1603
Db 742 KSIPTAS-----FGARPPVWRKIAVRAAOVLYNNKGYHSM 778
Qy 1604 SFLVNNAILKLANLOKE--NPSHYGTAFNHPNLTKOOLEVALMTTSVDLVASLCY 1662
Db 779 TYLSNNAILRANLPSKGNPAAYXITLVNHPNKTASLS--LDVLLQCTDVVIAIFLI 837
Qy 1663 FAMSFPASFPVFLIOERVSXAKHLOFISGVKPYIWLNSFWDMKNNYVPAVLIIPI 1722
Db 838 VAMSEFPASFPVFLVAKSKAKHLOFVSGCNFVITWLANVYVMDLNLVPAVTCVILDF 897
Qy 1723 CEQKSVYSSINLPVALLLLTGWSITPLMPASFVFIPTASTAVVLTSMVLTGINS 1782
Db 898 VFDPATFTSPNFPAYLSLFLGWSITPLMPASFVFIPTASTAVVLTSMVLTGINS 1956
Qy 1783 VAFVLELFT--DNKLNINILKSVFLPHHCISGLGLDMYKKNQAMADALERFGE--NRF 1840
Db 958 VAFVLELFT--DNKLNINILKSVFLPHHCISGLGLDMYKKNQAMADALERFGE--NRF 1840
Qy 1841 VSPSLMDVLCRNLFAMAVECEVFFLTVLLOYRFFIRPPVAKISPLINDECEVRRRQ 1900
Db 1018 KSPFEMDIYTRGLVAMTVGFGFELTIMCYQVNLQRPRLPVSKRPED--DVVASRQ 1906
Qy 1901 RILDGGGONDLEIKELTETIRRK--RKPAVDRICVGI--PPGECFGLLVGAGKSSST 1956
Db 1077 RYLKRGADNDMKIENLIVYVSKRTIGRLAVDRICLVGVGRCFGLLVGAGKSTF 1136
Qy 1957 KMLTGTITVTRDAPLNKSLISNIEHVNMGYCPQDAPTELTTELIGEHFEFFALLRGV 2016
Db 1137 KMLTGTITVTRDAPLNKSLISNIEHVNMGYCPQDAPTELTTELIGEHFEFFALLRGV 2016
Qy 2017 PEKEVKGCEMIRKLGIVKGEKAGYSGGNKRKSLTAMALIGPPVVELDEPTTGM 2076
Db 1197 PKKDAOVYKNALEKLELTETKADKPAGYSGGNKRKSLTAMALIGYPAFIFLDEPTTGM 1256

```

```

Qy 2077 PKARFLMNCALSVYKGRSVLTSHSMECEALCTRMALNVNGRRCGVSQHLKNRFG 2136
Db 1257 PKARFLMNLTLIDIKRGSRVVLTSHSMECEALCTRLALNVNGRRLGJCSIOHLKNRFG 1316
Qy 2137 DGYITVRIAGSNBCLKPVODFFGLAFPGSVLKEKRRNLDYQOLPSSLSLARIFSISQ 2196
Db 1317 DGYITVRI--TKSSQNVADVVRFFNRNPEAHAGCKTPYVQYQLKSEHLSLAQVSKMEQ 1375
Qy 2197 SKRRLHEDYSVSQTLDDQEVFNFAKQSD 2227
Db 1376 VGVGLGIEDYSVSQTLDDQEVFNFAKQSDN 1406

RESULT 4
probable ATP-binding cassette transporter ABC-3 - human
N:Alternate names: ATP-binding cassette transporter ABC-C
C:Species: Homo sapiens (man)
C:Date: 29-Jan-1998 #sequence_revision 06-Feb-1998 #text_change 02-Feb-2001
C:Accession: S71363
R:Klugbauer, N.; Hofmann, F.
FEBS Lett. 391, 61-65, 1996
A:Title: Primary structure of a novel ABC transporter with a chromosomal localization
A:Reference number: S71363; MUID:96326608; PMID:8706931
A:Accession: S71363
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-1704 <KUD>
A:Cross-references: EMBL:X97187; NID:91514529; PIRN:CA65825.1; P1D:e243436; P1D:9151
A:Experimental source: cell line medullary thyroid carcinoma
C:Genetics:
A:Gene: GDB:ABC3
A:Cross-references: GDB:3770735; OMIM:601615
A:Map position: 16p13.3-16p13.3
C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homolog
F:255-329/Domain: transmembrane #status predicted <TM1>
F:307-329/Domain: transmembrane #status predicted <TM2>
F:345-364/Domain: transmembrane #status predicted <TM3>
F:373-394/Domain: transmembrane #status predicted <TM4>
F:401-422/Domain: transmembrane #status predicted <TM5>
F:452-475/Domain: transmembrane #status predicted <TM6>
F:549-739/Domain: ATP-binding cassette homolog <ABC1>
F:566-573/Region: nucleotide-binding motif A (P-loop)
F:685-690/Region: nucleotide-binding motif B
F:1100-1120/Domain: transmembrane #status predicted <TM7>
F:1145-1169/Domain: transmembrane #status predicted <TM8>
F:1181-1207/Domain: transmembrane #status predicted <TM9>
F:1215-1236/Domain: transmembrane #status predicted <TM10>
F:1245-1264/Domain: transmembrane #status predicted <TM11>
F:1299-1324/Domain: transmembrane #status predicted <TM12>
F:1399-1590/Domain: ATP-binding cassette homolog <ABC>
F:1416-1423/Region: nucleotide-binding motif A (P-loop)
F:1535-1540/Region: nucleotide-binding motif B
F:674,866,1524/Binding site: phosphate (Ser) (covalent) (by cAMP-dependent kinase) #status pr
F:1344/Binding site: phosphate (Thr) (covalent) (by cAMP-dependent kinase) #status pr

Query Match 22.4%; Score 2645.5; DB 2; length 1704;
Best local similarity 35.7%; Pred. No. 1,4e-162;
Matches 639; Conservative 300; Mismatches 523; Indels 327; Gaps 48:

Qy 534 ACIVETGTPGSE-LPHNVYKIRMDIDNERTN-----KIKGYM----- 574
Db 131 MAUVFHFHNSKEPLPLAVKYHLF--STRNNYMTOTGSEFLKETGWTJTSFLPL 187
Qy 575 --DPPRADRPED--MKYVNGFAYLQDVVEQATIVLGTG-----KKTGYMOMPY 624
Db 168 FPNPSPRPPTSPDGEPEYIREGFLAVQHAVDRAIMYHADATROLQFLVITIRFPY 247
Qy 625 PCYVDLFLRVKSMPLFMTLAWISVAVIKGIYKARKEKRMIRINGDNLNLSMFS 684
Db 248 PPIADPFLVAIOYOLPLLLSTFTYALITAVAAVOEKERKLEKRYMRMGSLSWLHMSA 307

```

```

OY 685 WPISSILPLVLSAGLLVYLKLG-----NLBPDSVVEVFLSFAVATLLOCELLSTL 739
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 308 WFLFLPLFLILIASFMTLLFCVKRPNVAVLSRSDPLSLAFLCLFALSTISFSEFMSTF 367
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 740 FSNANLAAACGGIITYETLTPVLCAVADYGFPLKIFASLSPVAFCECEYFALFEE 799
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 368 FSNANMAAAGGLTYETLTYIFEFVAPRNNMTLSOKLCSLSSVAMAMGAQLIGKFEA 427
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 800 QGIGVOMDLFESPE-EDGFNLTISIMLFDPLTGYMTYIAVPGQYGIAPRYE 858
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 428 KGMGIGRDL-SPVNVDDFCGQVLMILLDSVYGLVTYMEAVEPFGQVQPYE 486
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 859 PCIKSYWFG-----ESDEKSHPGSNOKRMEPICEEPTHLKLGVSIGLNVKYYRDM 912
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 487 FIMSTYCGRPRAVAGKEEDSDP-----EKALRNEYFEAREDELVAIGIKHLKSYFVGN 543
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 913 K--VAVDGLALNFYEGQITSLCHNGAKTTMSITLGLFPPISGATYLLGDKISEMST 970
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 544 KDBAAVBDLNLNLYEGQITVLLGHNGAGKTTLSMLTGLFPPISGATYISGELISODMYQ 603
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 971 IROULGCPCHNVFEDMLTYEBHIFWYARLKLSEKHVAEMOMALDYGLPSKLSKXT 1030
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 604 IRKSLGICPOHDLFDMLTYAEBHLYFAQLKGLSKRCKREVEKOMLHIGL-EDKNRSR 662
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 1031 SOLSGMORKLVALAFAVGSKVYLDEPTAGVDPYSRRCITWELLKYRQRTIILSTHH 1090
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 663 RFLSGCMRKLSIGALIASGKVLILDEPTSGMAVISRRAIWMLLOROKSDRTYLTTHF 722
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 1091 MDEADVLDRLAITISHGLCCVGSFLKKNLGNIGYLLTVKKDVESLSSCRNSSSTVS 1150
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 723 MDEADLLDRLIAIMAKGLCCGSSFLPKQYKGAGYHMTLKE-----P 766
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 1151 YLKREDVSQSSSDAGLSDHESDYLTDVSAISNLIIRKHVSEARLEVDIGHELTYLPY 1210
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 767 HCNPD-----ISOLVHHNHPATLTSSAGAEISFLPR 800
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 1211 EAKEGAEVLFHEIDRLSLGSSISYISTETLEETFLKVAE--ESGVDAETSDGTLPA 1268
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 801 ESTHR--FEGFLPAKLEKQKOKELIASFGASITTEVEFLRGLKLVDSMDIOAIQ--LPA 856
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 1269 ---RRRRRFRG---DKOSCLRFETDDAADPND--SDIDESRETDLLSGMDKGSYQV 1319
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 857 LOYOHERRRSDMAVDSNLC-----GAMPDSDGALIEBERPAVKLNTCL----- 901
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 1320 KGWKLTOQOFVALLMKRLLIARRSRKGFPAQIVLPAVYCIALVPSLIVPFCKYPSLEL 1379
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 902 ---ALHCQGFMAFLKKAAYSREWKMYAAQVLYPLTCVTLAL----- 942
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 1380 QPMNYMEQTYVSNDAPEDTGLLELNLATKDPFGTKMEGNPIPTPCOAGEEWTTA 1439
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 943 -----AINYSSELPDPM--RLTLG-----EYKRT 966
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 1440 PVPQTTIMLDFONGNMTMNPSPACOSSDKIKMLPVCPCAGGLPPPOKRONTDILIOD 1499
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 967 VLFPSVGTSQLGOQLESHLKDQLAQEG-----QEPREVLD 1003
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 1500 LTGNISDYLVTKYVQILAKSLKNKIMVNEFRY---GGFSLGVSNTQALPSSQEVNDAI 1555
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1004 L-----EEFLI-----FRASVGGGFN----- 1020
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 1556 KOMKKHLAKKSSADRFLNSLGRFMTGLDTRNNVKNVFNNGKWAHAISSFLVANNALIR 1615
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1021 -----ERCL--VAASPRDVGERTVNALFNNOAHSPATALAVVUNLJFK 1063
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 1616 ANLOKGENPSHYGTIANHP-----LNLTKQQLSEVALMTTSVDVLYSCVIFAMSFVA 1670
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1064 --LLCG--PRASIVSNFQPSALQAAADQNE--GKGGIDALNL--LFMAAFILAS 1113
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 1671 SPVFLLOERYSKAHLQETSGVFPVLYLWLSNFWMCMNVVPATLVIIIFICEQOKSYV 1730
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1114 TFSILASERAVOKHQVSGVIVASFWLSTALLMIDLISFLPSLLLVFKAFDVARFT 1173
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

OY 1731 SSTNLPLVALLLLTGWSITPLPMYPAEFKIPSTAAVVLTSVNLFIGINGSVATVLEL 1790
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1174 RGHMADTLLLLLTGVALILPLMTLMNFPLCAATATATRLTFLNLSGI---ATFLMVT 1229
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 1791 ---FTDNKLNININDILKSVFLLFPHFCGLRGLIDMVK-----QAMADALEHFG- 1836
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1230 IMRIPAVKILIELSKTLDHVEFLVLPNHGICGMVSSFEYENETRRYCTSEVAHAHCKRYNI 1289
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 1837 ---ENRFVSPISMDL--VGRNLFMAVSEVPELFTVLYOYRFFIRPAPVAKL----- 1885
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1290 QYOENFY---AWSAPGVGFVAMASGAYALLFLLETNLORLGLICALRRRTL 1345
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 1886 -----SPUNDEDEYARRRORILDDGGONDI---LEIKELTKYRRRRK--AVDRICV 1934
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1346 TELYTRKRPVLEPDQDVADDETRLIAPSPSLHTPLIIEKLSKVY--EORVPLVADRSL 1404
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 1935 GIPPECEFGLLVGNAGKSSTFKMLTGDTTVRGCAFLLKNLSILNIEHYONMGCPOF 1994
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1405 AVOKGECFGLLGNAGKTTFMKLTVGESLTSGDAPFVGGHLSIDVCKVRQRTICPOF 1464
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 1995 DATTELLTGREHEFFALLRGVPEKGVGEMAIKRLGLVYKGEKYAGNIGSGKKRKL 2054
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1465 DALDHMTGEMLYMVARLGRIPERHIGACVENTLRLGLEPHANKLVRTYSGKKRKL 1524
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 2055 TAMALIGPPVYFLDEPTGMDRKARFLMNCALSVYKGRSVYLTSHSMECEALCTPM 2114
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1525 TGIALIGPAVIFLEDPSTGMDPVARRLLMDTVARARESKALITLISHMECEALCTPL 1584
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 2115 AITMNGRFRGCGSVQHKNRFGDGYTIVRI--AGSNPLDKPVDFGLAFPGSVLKEKH 2172
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1585 AIMOOGFCRGLSGFQHLKSKRGSYSILRAVQSEGOEALIEERKARVLDLFPQSVLEDEH 1644
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 2173 RMLDYOQLPSSLSLARIFTSLISQSKRHLIEDYVSQTTLDQVFNFA 2221
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1645 QGMVHYHLPGRDLISWAKVGIIEKAKERYGVDDYSOISLEQVLSFA 1693
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

RESULT 5

A59188

ATP-binding cassette transporter ABC3 - human

C:Species: Homo sapiens (man)

C:Date: 18-Feb-2000 #sequence_revision 18-Feb-2000 #text_change 17-May-2002

C:Accession: A59188

R:Comments: T.D.: van Raay, T.J.; Petry, L.R.; Klinger, K.W.; Landes, G.M.; Burn, T.C. Genomics 39, 231-234, 1997

A:Title: The cloning of a human ABC gene (ABC3) mapping to chromosome 16p13.3.

A:Reference number: A59188; MUID:97179225; PMID:9027511

A:Accession: A59188

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-1704 <CON>

A:Cross-References: GB:U078735; NID:q1699037; PIDN:AAC50967.1; PID:q1699038

C:Genetics:

A:Gene: CDB:ABC3

A:Cross-References: GDB:3770735; OMIM:601615

A:Map position: 16p13.3-16p13.3

C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homolog

Query Match 22.4% Score 2642.5; DB 2; Length 1704;

Best Local Similarity 35.7% Pred. No. 2-2e-162; Indels 327; Gaps 48;

Matches 639; Conservative 300; Mismatches 523;

```

OY 534 AGVETGITPGSIE-LPHNVKIKRMDINVERTN-----KIKDGYW----- 574
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 131 AAVVFEPFHSHSKRPLAVKYLRF---SYTRBNYMTQTGSSFLKETEGHNTSLRPL 187
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 575 --DPGRADPED--MRVWGFATLQDVVEBALIRVLGTGE-----KKTGVYOMQPY 624
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 188 FPNPGPRELISPDGGEPPYLRGEGFLAVQHAVDRALEYNADAATROLFORLTVTTKRPY 247
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 625 PCYVDITFLRWKSMSPLEMTLAMIYSVAVILKIGIYERKARLEKMTIMGIDNSTLWFS 684
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 248 PPIADPELVAIQYQPLLLSLSTTYALTITARAIVQEKERRLKEVRRMGLSSWLHMSA 307
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

QY 685 WEISSLELLVAGLLVYLKLG-----NLPPSDPSVVFVFLSVFAVVTLLQCELI STL 739
Db 308 WFLLEFLFLILIASFWTLLEFCVKRPNVAVLSRSDPSLVLAFLLCPALSTISSEFMSTF 367
QY 740 FSRANLAACGGILYFTYLLPVVLCVAMQDVVGFILKFFALSLPVAFGCGEYFALFEE 799
Db 368 FSKANMAAFGGELFFFYIIPFYAPRYRNMTLSOKLSCSLSNVAMAMAOILGKFEA 427
QY 800 OGIGVOMNLFPSPE-EDGFNTLTSISMFLDFELVGVMTWYEAFFPGOYGIRPMYF 858
Db 428 KMGVOMNLDL-SPVNVDDDECFFQVIGMLLDSVLGLVWYEAFFPGOYFPGOWYF 486
QY 859 PCKSYWGE-----ESDEKSHPGSNOKRMSSEICMEEEFTHLKLVSITONLYKVRIDGM 912
Db 487 FIMPSTWCKPRVAVAGKKEEDSDP---EKALRNTFYFAEPEEDVAGIKIKHLSKVRVSN 543
QY 913 K--VAVDGLANFTEGQITTSFLGNGAGKTTTMSILNGLPPTSGTAYILGKIREMST 970
Db 544 KDRVAVRDLNINLEYGOITVLLGNGAGKTTTSMLTGLPPTSGRAYISGYEISQDMVO 603
QY 971 IRONLGVCPQHNVLJEDMLTYEBHTWFAARLKGISEKIVKAKEMQMALDVGDPSSKLKXT 1030
Db 604 IRKSLGICPOHDLFDMLTYAEHLYTAQLKGLSKQKCEYVQMHLIIGL-EDKWNRS 662
QY 1031 SOLSGMQRKLSVALAFVGGSKYVILDEPTAGVDPYSRGIMVELLKYRQGTITLSTH 1090
Db 663 RFLSGMQRKLSIGIALIAGSKYVILDEPTAGVDPYSRGIMVELLKYRQGTITLSTH 722
QY 1091 MDEADVJGDRILATISHGKLCVSSIFLNOLGTGYLLVKKDVESLSGRNSSSTVS 1150
Db 723 MDEADVJGDRILATISHGKLCVSSIFLNOLGTGYLLVKKDVESLSGRNSSSTVS 1150
QY 1151 YLKKEDSVSSSSDAGLGSDESIDLTDVSAISNLRKRVSEARLVEDIGHETLYVLPY 1210
Db 767 HCNED-----ISQVHHVHNALLESAGNELSFILPR 800
QY 1211 EAAKEGAFVLEFHEIDRLSDLGSSYGSITTEIFLKAEE--ESSVDAETSDGTLPA 1268
Db 801 ESTHR--FESLFAKLEKKEKELIASFGASTITMEVEFLRKGKLVDSMDIQALQ--LPA 856
QY 1269 ---RRRRKAFG---DKQSLRPFTEDDAADPND---SDLEPSRETFDLSMDQKSGVQV 1319
Db 857 LQYHERRRRSDMAVDNSNLC-----GAMPDSDIGALIEERRKAVKNTL----- 901
QY 1320 KGMKLTQOQFVALLMKRLILARRSKGFFAQTIVRAVVCALVSLVPPGKYPSIEL 1379
Db 902 ---ALHCOQFMAMFLKKAAYISREKMYAAQVLYPLTCTVLAL----- 942
QY 1380 QPMWNTNEQVTEVSNDAPEDTGLLELNAITFYDPGFTCMGKNIPDTPQAGSEEMTTA 1439
Db 943 -----AINVSELPDDPML--RLTIG----- 966
QY 1440 PVPQITMDLFONGMNTMNPSPACQSSDKKLKMLPVCPGAGGLPPQKQNTADILQD 1499
Db 967 VVPEFVPGTSQGLGOLSHLKDALQAE-----QEPREVLGD 1003
QY 1500 LTRNITSDYLVKTYVQITAKSLKNKIMWNERY---GGFSLGVSNTQALPSPQEVNDAL 1555
Db 1004 L-----EEFL-----FRASVEGGGN----- 1020
QY 1556 KOMKHLKLANDSSADRFNLNGLFMTGLDTRNNVKNYFNNKGMWHAISSEFLANNAILR 1615
Db 1021 -----ERCL--VAASFQDVGERTVNNALFNQAYHSBPATIALAVVDMLLK 1063
QY 1616 ANLQGENSHYGTITAFNHP-----LNLTKQOISEVALMTTSDVLSVCIFFANSEVPA 1670
Db 1064 ---LTCG---PHASIVSNFPQPSALQAAQDQNE---GRKGFDALNL--LFAAFILAS 1113
QY 1671 SFVVFLLQERYSAKHLQITISGVKPVYIYWLNSFVDMQVNYVPATVILITFCFOQKSV 1730
Db 1114 TFSILAVSERAVQAKHVQSVGVHVASFWMLSLMLDLISFLIPSLLLLVFFRAFDVRAFT 1173

```

```

QY 1731 SSTNLPVIALLLIXGWSITPLMYPASFVKIPSTAYAVLTSLVNLFTINGSVATFVEL 1790
Db 1174 RQGNHADLLILLILIXGMAITPLMINEFFLGATATRLIRINLSGT-----ATFLMVT 1229
QY 1791 ---FTDNKLNININDIKSVLIPPHFLGRLIDMKN-----QAMADALERFG- 1836
Db 1230 IMRITAVLEELSKTLDFVLPVLPNHLGMAVSEVEYENETRRYCTSSVEAAHYCKYNI 1289
QY 1837 ---ENREVSPLSMDL--VGRNLFMAVEGVFFLITVLYQREFIPRPVANKL----- 1885
Db 1290 QYQENFY-----AMSAFCVGRFVMAASGAYILLLFLIETNLLQRLGILARRRRL 1345
QY 1886 -----SPUNDEDVAREORLIDGGQNDT---LEIKELIKYRRKKP--AVDRICV 1934
Db 1346 TELTRMPVLEPDDVADERTRILASPSDLHTPLIKELISVY-EGRPVLAVDRSL 1404
QY 1935 GTPGECFGLGVNAGKSSFTMLTGTVTYTRGDAFLKNSILSIHIVHQMCKYQOF 1994
Db 1405 AVQKGEFCGLIFGAGKTTTFKMLGSESLTSGDAFVGHRISSDVGVKVRQIGYQOF 1464
QY 1995 DATFELTGREHVEFALLRGVPEKGVGVGEMAIRKGLVYKERYAGNTSGNKRRLS 2054
Db 1465 DALIDHMTGKEMLVYARLRGIPERHIGACVENTLRLGLLEPHANKLVRTYSGNKRRLS 1524
QY 2055 TAMALIGPPEVLEDEPTGMDPKARFLMNCALSVYKGRSVYLTSHSMECEALCTRM 2114
Db 1525 TGIALIGEPVAVIFLDEPSTGMDPVARRLMDTVARARSGKAITTSHSMECEALCTRL 1584
QY 2115 AINWNRFRCLGVSQVHLKRRFGDGYTVIRI--AGSNBPLKRVQDFGLAPGSVLKEKH 2172
Db 1585 AINWQGFQCLGSPHLLKRRFGSSYSLRAKYQSEQCEALFEFAFYDLPPGCVLDEH 1644
QY 2173 RNNLYQVLPSSLSLARTISLSQSKRRLIEDYVSQTTIDQVFNVA 2221
Db 1645 QGVVHYHLRGRDLSMAKKVGLLEKAKKGYVDIVSOLISLQVFLSPA 1693

RESULT 6
T33783
hypothetical protein Y398C.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 17-Mar-2000
C:Accession: T33783
R:Becker, M.; Graves, T.; Yoakum, M.
submitted to the EMBL Data Library, October 1998
A:Description: The sequence of C. elegans cosmid Y398C.
A:Reference number: 421408
A:Accession: T33783
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1802 <BEC>
A:Cross-references: EMBL:AF101313; PDB:AC69223.1; GSPDB:G000023; CESP:Y398C.1
A:Experimental source: strain Bristol N2; clone Y398C
C:Genetics:
A:Gene: CESP:Y398C.1
A:Map position: 5
A:Introns: 45/3; 114/1; 195/1; 230/3; 543/3; 794/1; 849/1; 1036/2; 1099/1; 1132/3; 11
C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homolog
Query Match 17.5%; Score 2060; Db 2; Length 1802;
Best Local Similarity 27.9%; Pred. No. 1,6e-124;
Matches 551; Conservative 348; Mismatches 630; Indels 454; Gaps 52;
QY 385 YTPDTPATQVMAEVNKTQPELAVFHDLGSMDELSPKWTMTMENSQEDMLVRLMDSRD 444
Db 119 YAPWTATKQIMKIQNRY-----TANDL-----LNP-VAAVILKGLVNTAVPVLNTNMT 167
QY 445 NDHFWEQDLGDLDWTAODIYAVLAKHPEDVQSSNGSVYTRAFETNQALRTISRFEAC 504
Db 168 YKGF-----TTBEHMSWMOGQOS-----EC 189
QY 505 VNLNKLPIATEVWILNKSMLDEKRRFAGIVFTGISTGLPHVKKYKIRMDINVE 564

```


Db 190 DN-----PLL-----AGIVDDSLAKDKNPKDRFTYTRLNTH 225
 Qy 565 RNNKIKDGDWGPRADEPMKRYWG-----GEAY 595
 Db 226 RNSR-----NAFGDSYPMDSVSTRAYVYSGPINPDNDGSGPYWQEGFMT 273
 Qy 596 LODVBOAIRVLTGTEKK-----TGVMQOMPYPYVDDIFLRYMSHMLFMTLAMI 650
 Db 274 VQAVAVALTTEITGEDAQLTELLDSYQVSRPFPYSTKI-IEIGAFMVAIVIFSEMT 332
 Qy 651 SVAVITKGIYKEEAKLKTETMRIMGDINSILMFSEISLPLVYASGLVAILKGNIL 710
 Db 333 SVYIYRAVAYEKEEDLKYMRMGISQFINVAHFIIINAKLTFAYVYLIIIMHFAL- 391
 Qy 711 PYSDPSVEFVETSVFAVVTILQCFILSTLSRANLAACGIIYFTYL----- 759
 Db 392 -KSDMLTFELMIAFVYVYAFPMISSEPMNSATSATLISVFMMLLYFWYAFSSIDOT 450
 Qy 760 -PYVLCAVAMODYGFLLKTFASLSP-VAFGCEYFALFEEOGIGYOMDNLFESEVED 817
 Db 451 NPYPL-----GYRL-----INCINDIALNGLQLLAAYEQADGLKMGELFTPPSDN 499
 Qy 818 GFNTTSSIMMLPDTFLYGVMTYIEAVFPGQYGP-PRWYFPCYKSYWF--GESEDEK 873
 Db 500 NLTFGHALIYDGIIMILTYIEAVIPGSGVPOKPMF-VLPYWFPMSSGSKYDS 558
 Qy 874 SHPSNOKRMSSEICMEEPETHKLVSIONLVKYV-----RQMKYAVDGL 919
 Db 559 SDQOQOIQYADHYKLEKEPTDLIPTINVNLTKTYGTFEKKLFEDCKFKSGEKKRAYSNL 618
 Qy 920 ALNFEQGISFLGNAGAKTTMSLTGTFPPTSGATYALGDIKSEMTIRONGVCP 979
 Db 619 NLKWPQGCYLLGNHAGSKSTFSMLTGAASSGSAIYNDITSLPKIRREGLOCP 678
 Qy 980 QHNVLFDMLTVEEHITWEYARLKLGS-EKHVKAMEQMALDVLSPSSKLSKTSQLSGM 1037
 Db 679 QYNLFGMTYMEHLFEFAKLEKERTWDPPEARETIARLIDF-----KAFPMAGALSQGO 733
 Qy 1038 QRLISVALAFVGSKVYILDEPTAGVDPYRSRGIMWELLKYYROGRTIISTHMDADV 1097
 Db 734 KRKLIALALITGSEVWMLDEPTSGMDPGARHETWTLQKEKERITLTLTHMEERDL 793
 Qy 1098 GBRILIIHSGKLCGSSSLFLKNOIGTYLLTKKDVESLSSCRNSSSTVSYLKKEDS 1157
 Db 794 GBRILIAHGOLECCGSMFLKQOYGOGYHLFIY----- 827
 Qy 1158 VSQSSSDAGLGSDBESDTLTDVSAISNLKHYSEARLYVDIGHETLYLPEAKEGA 1217
 Db 828 -----YDTSTPDVSKTDTIIREYIPEAHVESYIGQENTYLL-SATHRPI 871
 Qy 1218 FVELEHETIDRLSDGSSYSGETLEIFLVAE-----ESGVDAETSDGLPAR 1269
 Db 872 FPLELEEDHOTQCGITISFGVSTTMEVFLKVGHTADERYNEHGIENDISE----- 925
 Qy 1270 RNRRAFGDKQSLRPTEDDAADPNDSDTDPESRETDLLSGMGKSGYQVKGMLTQOOF 1329
 Db 926 -----MIEKD-DPILODL-----RAQRYVGFILQMOHA 953
 Qy 1330 VALMKRLIARSRKGFPAQOVLPAVFCIALVSLIVPPECKYPSLELOPMWNEOYT 1389
 Db 954 KAFYERAIFFEFKWTQFPLQVFPVAYIVLWFTSOVLPSVKE----- 997
 Qy 1390 FVSNDAPEDTGLLELINALTKDPGFGTRCMGNPIPDPCQAGEEWTAAVPOITIMLF 1449
 Db 998 -----QDPQITSLA-----PFSDFK-KAG-----HLVSD- 1020
 Qy 1450 QNGNMTPONDSPPACQSSDKIKKMLPVCPPGAGLPPPRKQKNTADIIQDLTGRNISDY 1509
 Db 1021 -SGNYVTL-----LGG-----SOMLSMV 1038
 Qy 1510 VKTYVOALIAKLNKIMVNEFRYGSGLGVSNQALPQSOEVNDAIKOMKHLKIAKDS 1569
 Db 1039 QGIVTQ-----LGVYOT-----VVDITSNVEKFTIMQTNAM 1069

Qy 1570 ADR-----FLNSLGRFMTGLDTRNNKVMENKGMHAISSEFLVNNAILRANLOK 1620
 Db 1070 GSRFTGLHAYALCFVSSMNFST--VSVPSLKTFEPPNNGGLTALAITFTDSMILSQOK 1127
 Qy 1621 GENPSHYGTAEHNPBLNTRQOLSEVALMTTSVDVLYSICVIFAMSVVPSVFFLIQER 1680
 Db 1128 -----QSFYAVNHLLPFPSTODTLKNTNRSDGAFLIAYGLIVSFAVCVAGVSOFLTER 1182
 Qy 1681 VSRKHLQFISGVKVIYVWLSFVMDMGNVVPATVLIIFICEQOKSYSSSTN-LPVL 1739
 Db 1183 KRKSHMOLLSGIRPMWMLTAFIWDAAFPYRILCFPAIFITININATYHDFCVMLIT 1242
 Qy 1740 LLLILGWSITFLPMPASVFKIPSTAYVLTYSNLTGINGSVATVLELFTONKLN 1799
 Db 1243 LSFLLKGMALPPTVYFQCFESAPRGFMVYMHILGMSGIAPVLIISOTSSLDAGYL 1302
 Qy 1800 NDILKSVELIFPHFCLGRILDMKVNQAMADLEF-----GENRFS 1842
 Db 1303 WSIIFA--WLFPTYNISQIATVTFQENENYRIACKLDICTIPMKFAYVTCGCTASERLYVD 1360
 Qy 1843 PLSMIDLVRN-----LFMAVEGVVFLITVLIQYRFP-----IRPRYNAKLSPLND 1890
 Db 1361 NVLF--VGNRKGLVYVIFLAVOGFIYMWIMRENDQTKLFALIRCKADNPIMDITD 1418
 Qy 1891 -----EDEDYRKR--QRILDGQONDILKELTKTYRKRKRPAYDRICVGI 1938
 Db 1419 TDKVDERVDSDVIAEKSVYQRL--ANNKTAIVSNVNLVWYGNFN--AVKGVFHNVS 1474
 Qy 1939 GECEGLGVNCAKSSYFRKLLTGDTTGTGDAFLNLSILSNHEHYOMGCPQOFAT 1998
 Db 1475 KDCPGLGVNCAKSTYFQMLTGENTSSGDATVYNGVSKNMRREGANTGCPQYDAI 1534
 Qy 1999 ELTGREHVEFALLRGVPEKEVKGCEMAIRKLGLVYGEKAYNSGKNRKLISTAMA 2058
 Db 1535 KEMSGEETIWFARICRIPKEDIPKVNNAVIAIGIMYASROIKYSGNRRRLSLG 1594
 Qy 2059 LIGGPVFLDEPTTGMDPKRARRFLMCAISVYKEGSAVLTSHSMECEALCTRAIMV 2118
 Db 1595 IVGLPVDVILDEPTSGVDPKARRIIMILNRDLGLALVLTSHSMECEALCTRAIMV 1654
 Qy 2119 NGRFCLGSVOHLKRNFGDGYTIVRIAGSNPLKPYQDFEGLAFPGSVLKEKRNMLQY 2178
 Db 1655 YKFCFYSCCHISRSYGSTLLIRLKNR-DAEKTKSTKQTFGSGVIAKEHVIQLMF 1713
 Qy 2179 QLPSSLSLARIPTLSQSKRLHIEDYSQSTPLDQVFNPAKMD--GSDDD 2228
 Db 1714 DIPRGDSMSRIFEXLETVSTLSMWDYSLSQTLLEQVFIERSHAGVSSDSE 1766

RESULT 7

A:Probable ABC transporter (imported) - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: A84845
 R:Lin.X.: Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
 euss, D.; Niernan, W.C.; White, O.; Elsen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MUID:20083487; PMID:10617197
 A:Accession: A84845
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1816 <STO>
 A:Cross-references: GB:AE002093; NID:96598351; PIDN:AAC02761.2; GSPDB:GN00139
 C:Genetics:
 A:Gene: AE2941700
 A:Map position: 2

Query Match 16.8%; Score 1984; DB 2; Length 1816;
 Best Local Similarity 28.8%; Pred. No. 1.4e-119;

Matches 584; Conservative 282; Mismatches 592; Indels 572; Gaps 60;

QY 489 NENNOAIRIS-RFMCVNLKL-EPLATEVWLKSMELLDERKFAIVFTGTPGS 545
 Db 59 DETNNMIDILSLKFPDLRLVTKIFKDDIELETYITSANHYCSEVRNCSNKIGAVVH 118
 QY 546 IELPHVVKIKIM-----DIDNERINKIKDGVWDGPRADPE-----DMRY 588
 Db 119 EGGPHLFQDSIRLNTMAFAGFPNKSIMDN-----GPIYINDLEMOINTIPMOY 169
 QY 589 VAGGFAVLODVDEAIIHVLGTGTEKTSVYMOQMPYVDVLDILRMSRMLPMTLAW 648
 Db 170 SEGGELTLOOVVDSFII-----FASQON-----NDPLHSNLSALREFLPW 212
 QY 649 IYSVAIIIGIYKEAKLKETMRIMGIDNSIIMFSFISLLPLVLSAGILVYLKLN 708
 Db 213 TLFSPSVIRAMPFPFRTTDE-----FOSIVKSVMLF-----246
 QY 709 LIPYSDPSVVFVLSVFAVNTILOCFILSLFSRANLAAAGGIYFTLYLVLCVANO 768
 Db 247 LFKYSDKTLVFTYFELFLSLAIMLSFMISTFTTAKTAAVAGTILPLGAFPPY--TVND 304
 QY 769 DYVGFTLKIFASLLSPAFAFGCEYFALFEBOGIGVQWMDLFSPEEDGFNLJTSISM 828
 Db 305 ESYSVAVLKVAASLSPFAFGSTINFADYERAHVGLRMSNIMPA-----SSGSFVCLLM 361
 QY 829 LFDFTLYGVMTWYIEAVFGQYGIIPRWYFPTKSYWFGESDEKSH--PG-----877
 Db 362 LIDSTLYCALGLYLDKYLPRENGVRYPMNFTSK--YFGKKNNIIONRLPGETDFFPAD 419
 QY 878 -----SNOKRMSTI-----CMEEPTHIKLGVSIONLVKY--RDGKKVAV 916
 Db 420 IENVGEPFDPVRESISLEMKOELDRC-----IOVRNLHKYVASRRGCCAV 468
 QY 917 DGLALNFESQITSFLGNGAGKTTMSILTLGPPTSGTAATIGDISESTIRONIG 976
 Db 469 NSTDLTYENOMIISLGHNGAGKSTIISMVGLPPTSGDALILGNSTITNDELKREIG 528
 QY 977 VCPQHNVLEFMTLVEEHIWYALKLSERKAKENOMALDVGFSKLSKTSQLSGG 1036
 Db 529 VCPQHDLPELTVREHLEFAVLKGVESGSLKSTVVDMAFEVGL-SPKINTLVALSGG 587
 QY 1037 MOKRLVALAFVGSKVLLIDEPYTAGVDYSRGIMELLKTRROGRTIILSTHNDADV 1096
 Db 588 MKRRLSIGLIGNSKVILLDEPTSGMDPYSNMLTMOILKTKRITILLTHSDAEAE 647
 QY 1097 LGRRIALISGKLCCVGSILFLKNOLGTGYLLTVAKDVESLSCSRNSSTVSYLKED 1156
 Db 648 LGRRIIMANGSLKCCGSIFLKHNGVGYTLTVK-----TSPT-----687
 QY 1157 SVSQSSDAGLSDHESDTLIDVSAISMLIRKHSEARLYVEDIGHETLYVLYPEAKEG 1216
 Db 688 -----VSAAHIVHRHPSATCVSEVGENESFELP--LASLP 722
 QY 1217 AVEVLEHEIDRL---SDL-GISSYGISFTTLEIEFLKVAEESGVADETSDGLPARRN 1271
 Db 723 CEENMERLEIESCMKNSDSPGICISYVTLLEVEVLRA-GCNLDIE-----770
 QY 1272 RAFGGKOSCLRPTEDAADPNDSIDPRESREDDLSGMDKGSYVYK-----GW 1322
 Db 771 -----DKO-----EDIEVSPDKSSLYCIGSNOKSMOPKLIASCNDAAGV 811
 QY 1323 KLTQ-----QOFVALLM-----KRLIARRSKGF 1347
 Db 812 ITTSYAKAFRLIAVAVMTLLIGFISIOCGCSIIISRMFWHRKCALFKTRRSACRODKTY 871
 QY 1348 FAQIYLPVAFVCIALVSELYPPFGKTPSLELOPMATNEOYTFVNSDAPEDCTLELNA 1407
 Db 872 AFQITIPAVFLFGLF-----LQTKP-----HPDQKSTITLTAYFNP 909
 QY 1408 LTKDPGFGSTRMCEGNPIPTPCAGFEEMWTAPVQPIINDLFONG-----NMTMGNPSP 1461
 Db 910 LLSGKG-----GGGPIIPFD-----LSVPIAKVAVOYIEGWTIOLRNTSYKFPNP 954

QY 1462 ACCGSSDKIKMLPVCPPGAGLPPPOKONTAIDLODITGRNISDYLVKYVOIIAKSL 1521
 Db 955 -----KEALDAI-DAGPILGPTLLS-----975
 QY 1522 KNKIWNNEFRYGGFSLGSYNTQALPPSGEVNDALAKQKKHLKIAKSSADRLNLSIGREM 1581
 Db 976 -----MSFEIWSFD-----QSYOSSRE-----GLSSHSCNHPDGSIG---1009
 QY 1582 TGLDRNNVKKWVFNKGMHAISSEFLANYINNALIRANLOKGENSHYGIAPNPLUTK- 1640
 Db 1010 -----YVILHNGTOCHAGPIYINVMHALLR--LATGN--KNMTIOTRNHPLPPTKT 1057
 QY 1641 QOLSEVALMTSVLVVLSICVIFAMSVVPVAFVFLIOERVSQAKHLQFISGVKPYIWL 1700
 Db 1058 QIORHHDADFASAIIVNI-----AFSPFPAFVPIYKEREYKAKHQOLISGVSLSVWL 1113
 QY 1701 SNFVWDMKNVVPATLVIIIFICFOOKSYVSTNLPVLLIILLYGKSTIPLMPASVFE 1760
 Db 1114 STYVWDFISFLPSTFALIFAFGLFOFTIGIGRELPIYMLLEGLAIASTYCLTFE 1173
 QY 1761 KIPSTA-----YVULSVNLFT-----GINGSVAFVLELF--TDNKLNN 1798
 Db 1174 TEHMAQATSSYSVLPISLFVSPSSNVLLVHFPSSGLIIVISHWGLIPATASANSY 1233
 QY 1799 INDI-----LKSVELFHPFCIGRGLDM-VKNOMADALRRPGENRFPVSLDLYGR 1851
 Db 1234 LKELIFRYALQNFPRLSPECFSSDGLASIALIRQMKDKSH-----GVFEMWYGA 1286
 QY 1852 NLFMANEGVFFLITVLIQYRFFIRPVPNAKS-----GVFEMWYGA 1286
 Db 1287 SICYLGEVLECYCRYSMLLSFF--HGIDRLSLITVIGASRLTELIDRVYSTSFST 1343
 QY 1887 PL-----NDEVDREORIIDGGONDIELEIKTYRKR-----KAVVD 1930
 Db 1344 EPLIKSTGATSDMDDLDVOEENRVTSGISDMTILYOLKRVGRGDHNGRPKVAO 1403
 QY 1931 RIVCGIPRECEGGLGVNGAGKSTIFKMLTGPTVTRDARLKNLSIINIEHONMGY 1990
 Db 1404 SLFVSQAGECGFELTNGAGKTTLSMLSGHEPTSGTARIFGKDIVASPKALRQIHGY 1463
 QY 1991 CPOFDAITELTGREHVEFFALLRGVPERKGVKGEAMIRKLGVLKGYKAGYSGGNK 2050
 Db 1464 CPOFDALFELYLVKEHLEIYARIKVYDHRIDNVYTERKIVFEDLKSHPSPFLSGGNK 1523
 QY 2051 RKISTMALIGSPVVELDEPTTGMDPKARFLNN--CALSVKEGRSVYLTSHSPECE 2108
 Db 1524 RKLVAIALMIGDPPIVILDEPSTGMDPVAKRFMDVYSRLSTRSGKAVIILLTHSMNEAO 1583
 QY 2109 ALCTRAIIVNCRFPKGLCSYQHLKNRG-----2136
 Db 1584 ALCTRIGIIVNGRLKCTSPQHLKTRYGNNLELVEPPYNGVKNPVEYNVELNFCOIQQ 1643
 QY 2137 -----DGYT-----IVVRIA-----2146
 Db 1644 WLFNVPTOPRSLGLDEVICIGVDSITPDTASAEISLSEPMOIRAKFLGNORVSTLV 1703
 QY 2147 -----GSPN-----DLKVPDFFGLAFPGSVLKEKR 2173
 Db 1704 PPLPEDEVRFPDOLSEOLFRDGGITPLPIFAEWMLTKKFSALDSFIQSSPGATFSKNG 1763
 QY 2174 NMLQYOLFSSLS--SLARFISLSQSKRLHIEDYVSOTTLDOYVNVNA 2221
 Db 1764 LSIKYOQPFEGGGLSLADAFGHLENRNKRGLAIETYSISOSTLETTFNHRA 1813

RESULT 8
 T15200
 hypothetical protein F12B6.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Mar-2000
 C:Accession: T15200
 R:Pauley, A.; Maggi, L.

submitted to the EMBL Data Library, May 1997

A:Description: The sequence of C. elegans cosmid F12B6.

A:Reference number: 218307

A:Accession: F12B6

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1447 <P>AD

A:Cross-references: EMBL:AF003138; NID:g2088708; PID:g2088709; PIDN:AA54153.1; GSPDB:GN

A:Experimental source: strain Bristol N2; clone F12B6

C:Genetics:

A:Gene: CESP.F12B6.1

A:Map position: 1

A:Mutons: 79/2, 114/3, 177/1, 224/3, 331/1, 345/3, 373/2, 417/2, 464/1, 536/1, 659/2, 6

C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

Query Match 15.8%; Score 1863.5; DB: 2; Length 1447;

Best local similarity 28.3%; Pred. No. 6, 1e-112;

Matches 505; Conservative 283; Mismatches 498; Indels 499; Gaps 51;

571 DGYWDGPRADPEDMRKYGAFVLO-----DVEQAIIIVLTG-TEK 613

2 DSKNSPDRKPLDLKTIIFGFSFLOGSWPSEKSKQKSEIDRAIMSELTNOTDA 61

614 KTYVMOOMPYPCVDDIELVMSRNPFLMTAMTYSVAVITKGYVEKKAIRKETMRI 673

62 NLGYAAGEPPTVKTDF--NVALEMPFLLSIFPSPALKNIVYERKQIKQOMRA 119

674 MGLDNTIMFSPFISLPLVLSAGLVLTICNLIPYSDPSVVFVLSFAVVTIIQC 733

120 MGLDGVHVFISWGLISLVANFISLITISAKAFIDYDYLFLVILFLESSIAMS 179

734 FLISTLFSRANLAAAGGIIFFLYLPLVLCVAMQDYGF-TLKIFASLSPAFPGCG 791

180 IFFSLFLFNANLAPATCVLWFVFFIFPOLKRT--DRISPTENRISLILPAMGH-CP 236

792 ---EYFALFEQIGVQVNDLFESEVEDGENTLTSISMIFDFFLTGVMTVYIEAVFG 848

237 KLESEFNAMER---ATMSDLMEKNPVLGIVSELCMMLVVDVAVFLILAVYISAVAPG 292

849 OYGIKRPWFYPCPKSYW-----FGESEDEKSHPGSQKMSSEICMEFEPTNHLKG 898

293 DGVNOLPLFEPPTLKVMAPGLYKNVEFEVDDEHEDTIPNSOS-----FSEETNLTLL- 344

899 VSIQNLVKKYRGMKVAVDGLALNFYEGQITSPFGHNCAGKTTMSILTGLEPTSTGAY 958

345 ---ADCLNLRLYEGQITGLGHNAGKTTMSILGLEYABSSGTA 388

959 ILGKDIRSEMTIRONLGYVCPQHNVLFDMLTVEEHIMFYARLKLSEKHYKAEQDQALD 1018

389 IYORDIRDLRRVRYDLGICPQHNVLFSHLIVSEQLRLFAALKGVPDSDELTSQVDELLAS 448

1019 VGLPSSKLSKTSQSLSGNQKLSVALAFVGGSKYVILDEPTAGVDPYSPRQIGWELLKY 1078

449 VSL-TEKANKLASTISGSGRRICIGIATIGSRVYILDEPTAGVDPYSPRQIGWELLKY 1078

1079 ROGRTTILSTHMHDEADVLDGRLTAIISHGKLCVGSILFKNQOLGTGYLLTVKQDVSS 1138

508 KEGTITLSTHMHDEADVLDGRLTAIISHGKLCVGSILFKNQOLGTGYLLTVKQDVSS 1138

1139 LSSCRSSSTVSTYKKEEDSVSSQSSDAGLSHDSFTLLIDVSALSINLRKHVSEARLVE 1198

539 --- 538

1199 DIGHELYVLPYEAKEGAFVELFHEIDRLSDLGISYGISSETTLEIFLKVAESGVD 1258

539 --- 538

1259 AETSQGLTPARRNRFAQKQSCIRPTEDDADPNDSOIIDPESRETDLLSGMDGKSYQ 1318

539 ---KPDLL---DGR--- 546

1319 VKGKMLTQGOQFVALLMKRLILARRSKRQFPAQIVLPAVFCIALVPSLI--- 1367

Db 547 ----RLIQHFPALLVCRINVTLSKSKRTFLQVITPLFLLAELFVLQVSTARPDLNV 602

Oy 1368 -VPEPKYPSLELOPMWYNEQTYVSN-DAPEEDGTLELNLALTKDGFECTRCMEG--NP 1423

Db 603 SMPPLPLETSI-----MGNSHDFVNSWMTAENSTANDIILHMFSSFGTGPCKAKVDPND 657

Oy 1424 IDPT-----PCQ--AGEBEWT-----TAPV 1442

Db 658 LDBTMRRELMEFRNRFGFGKRRKAPGVDDSDVNEQCCNQIGGEDYEDTISNATYNAPIY 717

Oy 1443 QTIMDL-----FQNGWMTQNPSPACQSSDKIKKMLPVCPPGAGLPPRQKQNTFADIL 1497

Db 718 CGCEDGWNCTLEDMKMNEN-----WLRNTTDRI 748

Oy 1498 QDITGNISY-LVKYVQDILAKSLKNLWNEFRYGFSGVSTQALPSPQEVND-- 1553

Db 749 FDLTGRNLQFRLITFPQALNTA-----PFLGGFSLGHVQRA--QSOADIDRSK 799

Oy 1554 ----AIQMKKHLKL-----AKDSSADRLNSLGRMTQGLDPFNN 1589

Db 800 RGMLETIKDIAQSMRLINLNTTGIEPATPKVLDPPAQNTITLQVYVNDL---LQNLQVREN 856

Oy 1590 VKVFNKKGHIALISFLVNNALILRANLQKGNPSHYGTFANPLNLT-KQOLSEVAL 1648

Db 857 VKVFNKRIKPPPIASNLISNALLQD-DYALDPEDDGLIMNHNPKTISQTLQNAL 915

Oy 1649 MVTSDVAVLSICV-IFAMSFVPAFVFLIQEVRVSAKHLQFISGVKPYVTLNSFWDM 1707

Db 916 KFTQALAVFRITILLVLSMIPAGETVYIVEDRICALDQLQGGIRKQYVWVSTSLDM 975

Oy 1708 GNYVPATLVITITFCQOKSYASTNLRYLALLLGYGSTPLMYPAFPIKIPSTAY 1767

Db 976 VGGIHP-----RHICNNALPVLPL-CLRLYRRNRILALPSPIL----- 1013

Oy 1768 VLTGSVNLFIGINGSVATFVLELTQNKLNINDIL-----KSVFLIPPHKCLGRILDM 1822

Db 1014 ----RANVDLSLCLIPKSLFCGSGSLFCNCWFL 1046

Oy 1823 VKQAMAD---ALERFGENKRVSP-----LSMDLVGRNLFMAVEGVFFLTV 1868

Db 1047 RRRSLCLDSYHARRVAYGSEQNNRPDMINOLPLPSLAPDQGHIMCLFTHVILATICLI 1106

Oy 1869 LLOY-RE-FIRPRYNAK-----LSPLENDEEDYRERORI---LDGGQNDILEKEL 1917

Db 1107 FSDMEFGVRRKRENLDAMMLREPSRCDDEDVYKEROVDALPMD-SSDNHALIYRNL 1165

Oy 1918 TKIYRRKRPAYDRICVGIPECEGGLGVGAGKSSFEKMLT-----GPTVTRGDAFL 1972

Db 1166 AKAYNDEL-AVKGISFAVEPECEGGLGAGAGKTTFFAMLTAKIRPGHGSIMQNTRI 1224

Oy 1973 NKNLSLINSHEVHONNGYCPQDATTLELTGREGHEFPALLRGVPEKEVGKGMATIKL 2032

Db 1225 NTGS-FSDVNR-FQQLGCTGQFDALNMKSTRENKLFYARIGLIVPTQIDSLIRLLIAL 1282

Oy 2033 GLVYGEKAYAGNYSNGNKKRLSTANALJGPPVYLDDEPTGMDPKARFELNICALSVK 2092

Db 1283 HLRPYANTQTSLSGGRNRKRLAVALVSPSLIFLDEPSAGMDGSOQFLMKYIERLCK 1342

Oy 2093 EGSSVYLTSHSMECEALCTRNAIWNCGFRPLGVSQHLKKNFFGQGYTVIYRIAGSNPL 2152

Db 1343 SGRAVYLTSHSMECEALCTRIAIMRGRIICLIGGQHLKSKYKSGSMITKMG-GKDENA 1401

Oy 2153 KPYODFFGLAF-PGSVLKRNMLQYOLPSSLSLARFTLSQ 2196

Db 1402 KEIAGIMRSKLGDSRVEALHICSTIFIHIEQGTASVAVALEYVNO 1446

RESULT 9

C88925 protein F33E11.4 [imported] - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001

C:Accession: C88925

```

Db      621 GYTYQ-----LGYQT-----VVDITSNVEKFI---MDQT- 647
QY      1571 DREFLNSLGRMTGIDTFRNNKVMFNKNGMAISSPLNVLNNALLRANLQKGNPSHYGTT 1630
      1648 -----NMGSKRTFG-----HYALGFV-----PSMNFNS 671
QY      1631 AFNHPLNLTQOJSEVALMTSDVDTLVSJCVIFPAMSEVPASFEVFLIOBRVSAKHLQFI 1690
      672 TVSVF-----SLKISFAVCV-----AGSQFLITERKKKSHMOLL 707
QY      1691 SGVKPVITWLSNFWDMCNVVPATLVLIIIFCFOOKSVYSTN-LPVATLALLTGMSTI 1749
      708 SGIRPMMFWLTAEPIDAMAFVIRIICFDPAIETIFNTATTHDQGVALLITFSLFLYGWTA 1767
QY      1750 TPIMYASVPEKIPITPAIVVLSVNLFTGJNGSVATFEVLEFTDKLNNINDILKSVFLI 1809
      768 LFTYTFQFFESAPKGEWMTMYHILGTMGISIAVPIISOTSSLDAGYLSIIIFA-WL 825
QY      1810 PFHFCLGRGLDMVKNQAMADALERE-----GNKRVSPSLMDVGRN 1852
      826 PFTYNISQIATVTFQENVBVIRACKKLDCTIPMEKAVTACCGTASERLIVDVNLF-VGNR 883
QY      1853 -----LEAMAVEGVFFLITVLIQYRPF-----TRPRVNAKLSPLND-----E 1891
      884 KGIIVYVFLFVANOGHTIYIWMFMRENDQFTKLFLIRCKRADDNLTMDITTDKXDERHVE 943
QY      1892 DEDVRER---QRIIDGGGQNDILEIKELTYRKRKPAYDRICVGIIPGECGGLGVN 1948
      944 DSDVIAEKSVYQRL--ANNKKTALVSNNLVKMYGNFN--AVKGVNPHVNSKDCGGLGVN 999
QY      1949 GAGKSGPKMLTGTQTYTRGDAPLKNKSIIISNHEVHQNGCYCPQFADITELTGREHVE 2008
      1000 GAGKSTTFQMLTGNSTISSGDAYVNGSVNMMREAGANTGCPQYDAIIEKMSGECTLY 1059
QY      2009 FFALLRGVPEKEVGKVGEMARIKGLVYKGEKKAAGNSGGNKRLISLGMALLIGPPLYL 2068
      1060 MFARITRIPEKDIPKKNVAVIAHIGICMYASRQIKTYSGGNKRLISGLIAIVLPPVLL 1119
QY      2069 DEPTGGDPPKARIRIIMCALSUYKEGRSVYLITJSHMECECALCTRMAIYNGFRFCIGSV 2128
      1120 DEPTSGVDPPAKRKLIMVITLNRLLKDLGTALVLTSHSDECEALCTETELAIYNGKFCRYGSC 1179
QY      2129 QHLKNNREGDGYTIVRIAGSNPDLKPVODFFGLAPFGSVYKEKRRNMLOXYOPLSSISLSLA 2188
      1180 QHLKSRGSGSYTLILRLKLN--DAEKTASYIKQTFPGSVYKKEHVLQJLWFNDIPROGDSWS 1238
QY      2189 RIFSLSQSKRRLLHEDYVSVSTTLDQVFNPAKD---QSDDD 2228
      1239 RLFERKLETVSTLNDMDYLSQTTLEQVFIERSRQAGVSSDE 1281

```

RESULT 10
F88559
protein C48B4.4b [imported] - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 17-May-2002
C:Accession: F88559
R:Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode *C. elegans*: a platform for investigating bio
A:Reference number: A75000; MIMD:99065613; PMID:9851916
A:Note: see websites www.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C-elegans/
A:Accession: F88559
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1758 <STO>
A:Cross-references: GB:chr_III; PIDN:CAA82384.1; PLD:g3875025; GSPDB:GN00021; CESP:C4
C:Genetics:
A:Gene: C48B4.4b
A:Map position: 3
A:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homolog

Query Match 13.1% Score 1542; DB 21; Length 1758;
 Best Local Similarity 23.8% Pred. NO. 6.5e-91; Indels 558; Gaps 65;
 Matches 520; Conservative 371; Mismatches 740;
 162 LSPKSTVGMKLRADYILKRVFLQGLHLNLSGKSGSEEMTOL-GDOEVELGCLPE 220
 7 IIPPHSTIYRMLRRLYQNKAKRKAKESDIEGSSSESNSTPLRGDGLAESRIIPK 66
 221 KLAABER-----YLRN-----MDILKPLRLTSLNTPSPFSKLAELATYTLHSL 265
 67 KMRRLQFSLMLKDWVLLRRNKVWTLFELLPLCL-----LL 103
 266 GTLAQELFSRMSDMROEYMEFLTNYNSSSSSTOYQAVSRIVCGHPREGGLIKSLMWY 325
 104 GPLV-----YLVYKNADHTSSPENIY----- 124
 326 EDNNYKALFGNGSTEDAEETFYDNSTPYCNDLMLKLESPSRIL--WKALKPLLVGKI 383
 125 -DN-----FQVKGSTVEDY-----FLESNFKIPYKRCCLSDVAVG-- 159
 384 LYTPDTPATQVNAEVNKTQELAVFHDLGMEELSPKIVTEMENSQEMDLVRLDSR 443
 160 -YTSKDAARKTVDLKKFAE-----RQSAKILSYKNESSE--QLITVL 204
 444 DNDHFWEQDLGLDWTADIVAFLLAKHPEDVOSSNGSVYTWREAFNETMOAIRTSREME 503
 205 RND-----LPLMETFCALNS----- 220
 504 CVNLKLEPIATEVWLINKSMELDERKFWAGIVFTGIPGSI.ELDHVYKIRMDIDNV 563
 221 -----YAGVVFDEVDYTNKKL-----NYRLLOKTPRE 248
 564 ERTNKIKDWDPPGRADPFEDM---RYWGGFAYLDVDEQALIRVLTGTEKTKGVYA 619
 249 EWMHLETSYNPRGSSGRYSRIPSPRYWTSALFQOHAIIESFSSVOSGADPLPIL 308
 620 QOMPRPCYVD-----DIFLRVWSRSMPLMTLAMYSAVILIKGIVYEKKEARKETMR 672
 309 KCLPBRKRTSSVSATDF-----PRTYMAFVTFINVIHITRELAENHA-VKRYLT 359
 673 IINGLSILMFWSFTSLIPPLVSGALLVILKGLNLPYSDPSVFVFLSVFAVVTILQ 732
 360 ANGSLTFMYYAAHVMAELKFVI--FLCSITPLTFVMEFSPALLIVTVMYGLAVIF 417
 733 CPLISTESRANLAAACGGIITFYLYLVLCVAMODVYVGTALTFASL--LSPV----- 785
 418 GAFVASFENNNTSAIK-----ALIVAWGAMIGISYKILRELDIOSSCFIXGL 464
 786 ---AFGFCCEYFALFEEOGIGVQMDNLF-ESPVEDEGFNLTTISMLFDTFLYGVWTW 840
 465 NINGAFALAVEALSDYMRERELNLTNMFNDSSLH---FSIIGMALVMMIVIDLMLSIGAL 521
 841 YIEAV-FPGOYGIIPRPWYPTCKSTWFGESSEKSHPG--SNOKMSMLCEBEETHL-- 895
 522 VVDHRTSADFSLRTLDFE-----APEDENOTDGTVAQNTINQDMNPMASTINP 574
 896 -----KLGYSIONLVKYYRDKKVAVDVADIALNFEQGITSEFLG 933
 575 PNADSDSLEBSTDGACARDTARADITIVANLYKIMSTGERAVDGLSLRAVAGQCSILIG 634
 934 HNGAKGTTTMSILTLGEPPTSGTAYILGKDIRSEMSTIQNLGVCVPOHNVFLDMLTVEEH 993
 635 HNGAKGTTTSSJAGIIRPNGRITTCGYDVNEPGETRRHIGCMQVYVPLDOLTVSH 694
 994 IWFARAKTSGSEKHVAEMOMALDGLPSSKLSKTSQSLGSGMOKRISVALAFVGSKV 1053
 695 LKLVYGLKGRKEDOKODMKRLSLDYKL--DFKENEKAVNLSGMMKRLCVMALLGDSFV 753
 1054 VILDEPAGVDPYSRRIGWELLKRYGRTIILSTHMDADVLDGRIALISHGKLCVG 1113
 754 VLDEPAGMDPGARQOVQKLVEREKANRTIILTHYMDAERLIGDWVIMSGKCLVASG 813
 1114 SSLFLKNOGLGTGYLLTLVKKDVSSLSKSNSSSYVSYLKEDSVSQSSDAGLSDHBS 1173

814 TNOYLKQKFGTYLLTV-----LDHNG 836
 1174 DRLTIDVSAISMLIKRHVEARLVEDIGHELTVLYPYEAKEGAFELFHEID----- 1226
 837 DKRKANV-ILTVCHGHYVEAEREMHGOQIEIILPEARKKE--FVPLFOALEIODRNY 893
 1227 -----DRISDLGSSVGSISFTLEIFLKAEEGVDAAETSDGTLPARRNR 1273
 894 RSNVFDNMNTLKSQATLEMRSPGLSLNTLEOVITITGDK-----YDAAISRQNSR 946
 1274 AFGDKSCLRPFEDDAADPNDSIDIPRESRETDLLSGMDGSGSYOVKWKLTQOQFALL 1333
 947 TSNMNSNAPSLKPRAGYDTQSTKASDYCK--LMSQARCP-EKSGVARNVAOFISIM 1003
 1334 WKRLIARSRKGFPAQIYLPVAVFCIALVPSLIYPPGKYSLDLOVMYNEQYTFVSN 1393
 1004 RKKFLYSRRMAQLFQVILP--IILGLVGLT-----TLKSN 1040
 1394 DAPEDTGLELLNAITKDPQFGTRCEGNPIPDTPQAGEEEMTAPVPOITMDLPONGN 1453
 1041 NTDO-----FSVRSI-----TP--SCIE-----PSKYVMRFENGCT 1068
 1454 WTMONPSACQSSDKIKMLPVCPREGAGLPPPOKQNTADILQDITGRNISDYLVTRY 1513
 1069 -----IPEAANFEKILKRSGGFEVLNTYKNP 1096
 1514 VOITAKSLKNIWNEFRYGGFSLGVSNTQALPPSOEVDATIKOMKKHLAKDSSADRF 1573
 1097 LPNITKSL-----IGF-----MPA----- 1111
 1574 LNSLGRFMTGLDTRNNKYNVFNKNGHAISSFLNVLNAILRANLOKGNPSHYGTAFN 1633
 1112 --TIGMTMNS--DNLEALFNMRYYHVLPTLLSMINRAKLTQVAAELISSGVLYSKFT 1165
 1634 HPLNLTQOLESVALMTSYVDVLCVIFAMSPFVAFVFLIOBRVSKAKHLOFISCV 1693
 1166 SNSNLLPSOL-----IDVLAFLMLILFANVSTFVFWFLIEERTCOFAHQOFLGTI 1216
 1694 KPAIYMLSNFWMCMNVVPATLVIIIFICFOOKSYSSNLPVLAALLLLQWSTTPLM 1753
 1217 SPITFYSASLIYDGLIYSLICLIFLFLFAE--HMMYDHLAIVLFWLFFSSVPT 1272
 1754 YASVAFKIPRATYAVVLTSMVLEFINGSVAFVLELFTONKLNINDILKSVFL----- 1808
 1273 YAVSFLOSPSKANVLLIWOVVISGAALLVFLIFM-----IFNDWMLKSLIVNIFM 1327
 1809 IFPHFCIGRLIDMVKNOAMADALERFGENRFVSP---LSMDLVGRNLFAMAVEGVFF 1864
 1328 LIPSYAFGSAII-----TINTYG--MILPSEELAMWDHCGNAMLMGTFVCSF 1374
 1865 LITVLIQYFFLR-----PRYNAKLSPINDC--DEIVREBROIRILDGGGONDILE 1913
 1375 ALFVLLQFQFVRFLSQVTVBRSSHNNVOPWMMGGLPYCESYSEBREVHNVNQNALV 1434
 1914 IKELTKIVRRKKRPADVRLICVGIIPREGCFGLLVGNAGKSSTFKMLTGDITVTRGDAFLN 1973
 1435 IKDLITTF--GRITVANELCLAVDQKECGLLGVNACAKTTFNITIGOSPASSGEAMIG 1492
 1974 KNSITLSINEHVHONMCGYCPQFDALITELTGRBHVFPFALLRGVPEKEGVKGEAMIRKLG 2033
 1493 GRVDTTEI-----SICYCOFALMLDITLGRSTLELLOMGEFENYKA--KAELILECVG 1545
 2034 LVUYGEGYACNAGSGKKRKISTFAMALIGGPPVFLDEPTTGMDPKARFLMNCALSVKE 2093
 1546 MIAHADLVAFYFSGGOKRKISYVALLAPQMI ILDEPAGIDPAREVWELLMLMCREH 1605
 2094 GRS--VVLTSMECEALCTRNAIYVNGRFLCIGYQHLKNRFGGTYITVVIASNSDL 2152
 1606 SNSALMLTSHMDECEALCSRIAVLNKSLIALIGSSQELKSLYGNNTYTFLSLYEPNORD 1665
 2153 KPVYDFGLFAPGCVLKEKIRNM---LOYQDPSISLSS--ARFISILISQSKRLHIEDYSV 2208

Db 1666 MWVQ-LVQTRLPNSVLTSTNKTILNKKQIPKEKEDCSAKFEWQALANDGVKDFEIL 1724
 QY 2209 SQTTLDOVAVNA---KQSDDDHLKDL 2234
 Db 1725 AQSLEETFLRLAGLDEQDLDPHSTVEIS 1753
 RESULT 11
 T42749
 ATP-binding cassette transport protein homolog - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C:Date: 11-Jan-2000 #sequence-revision 11-Jan-2000 #text-change 21-Jul-2000
 C:Accession: T42749
 R:Wu, Y.C.; Horvitz, H.R.
 Cell 93, 951-960, 1998
 A>Title: The *C. elegans* cell corpse engulfment gene *ced-7* encodes a protein similar to A
 A:Reference number: Z22259; MUID:98297348; PMID:9635425
 A:Accession: T42749
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1704 <WUY>
 A:Cross-references: EMBL:AF049142; NID:93172340; PIDN:AAC24116.1; PID:93172341
 C:Genetics:
 A:Note: *ced-7*
 C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology
 Query Match 13.0%; Score 1538.5; DB 2; Length 1704;
 Best Local Similarity 24.9%; Pred. No. 1e-90;
 Matches 484; Conservative 347; Mismatches 675; Indels 441; Gaps 59;
 QY 458 WTADIV-----AFLAKHEDVQSSNGSY-----TWRAFNETN----- 492
 Db 24 WTLPELILPLLGLPLVLYLVKKNADHTSSPENLYDNFOYKGTVEDEVFLESNFKIPYKRW 83
 QY 493 -----QAIRTI-----SRFMECVNLNKL-----BPIATFVWLINKS 523
 Db 84 CLRSVVVGYSKDAAKRTVDLKKFAERFQSAKLKLSVKNESSEQLLT---VLKND 140
 QY 524 MELIDE-----RKNAGITFTIGTIGSIELPHHNYKTKRMIDNVERTNKTIDGYWDPG 578
 Db 141 LPLNLTFCALNSYAGVGVFDEVDTNKKL---NYRILKCTEPEWHILRETSYNPYGP 196
 QY 579 RADPEEDM---RYWVGFAVLDQVVEQALIFVLGTCKKGTGVMOCPRYCYVD----- 629
 Db 197 SSGKYSRIPSPSPWMTSAFLTRQHAIESSFLSSVQSGAPDLPILTRGLPERRYTSSVSA 256
 QY 630 --DIFLRVSRSMPLFPLNLAITYSAVLIKCIYVEKEARLKEIRINGLDSILMFSWFI 687
 Db 257 FLDFE-----PRTMAFVTETINVIHTREIAENHA-VKPYLTAMGLSTPMFYAHV 307
 QY 688 SSLILVLSAGLLVILKLGMLLPSPSVYFVLSVFAVTTILQCLISTLSRANLAA 747
 Db 308 MAFLEFVI--FLGSIILPTEVMEFSPAALITVLMIGAVLFGAFVAFVSPNNNSAL 365
 QY 748 ACGGIIITFLPLVLCVAMODYVGLTKIFASL--LSPV-----AFGCGEVEAL 796
 Db 366 K-----ALIVAMGAMIGISYKIRBELDQISGFIYGLNINCFALAVAIASD 412
 QY 797 FEEOGIGVQWNLF--ESPVEEDGFLNLTSSIMLFDFFLIGVMTWITIEAV--PPQYGIQR 854
 Db 413 YMRREBELNLMFNDSSILH--FSLGALVYMMIVDIMNSIALVVDHRTISADSLRT 469
 QY 855 PWYPPCKTSYWFGESEDEKSHPG-----SNOKMSEICMEEP---TJHL--- 895
 Db 470 LFDFE-----APEDEENQTDVTAQNTFRINDQVRNVRBSRSMELQMPMASTSLNPN 522
 QY 896 -----KLGVSNLVKYVRDGMKAVDGLALNLYEGQITSEGLHN 935
 Db 523 ADSNLSLEGSTEADGADTARADITVRNLIKWTSTGEKAVDGLSLRAVRCQCSILLGN 582
 QY 936 GAGTTMTSLTGLPPTSGTAVILKGRDIRSEMTIRQMGVCPHNVFLMDLVEEHIM 995
 Db 583 GAGKSTFSSIAIGIRPTNGRITICGYDVNGEGRIRIIGCMFOYNPLDOLTVSEHLK 642

QY 996 FYARKLGSSEKHVKAEMQALDVGLPSSKLKSTQSGMORKSLVALAFVGGSKYVI 1055
 Db 643 LVYGLKGRAREKDFQDMKRLILSDVKL--DFKENEAVALSGMKRKLCCVCAALIGDSEVL 701
 QY 1056 LDEPTAGVDPYSRRGIWELLKTYRQGRITILSTHMDADVLDRIAIISHGKLCVSS 1115
 Db 702 LDEPTAGMDPGARODVOKLVEREKANRTILTTHYDEKERLQDWFIHSHGLVASGTN 761
 QY 1116 LFLKNQLOCTGYTLVLRKDVESLSLSCRNSSIVSYLKKEDSVSSQSSNAGLSHESPT 1175
 Db 762 QYLKOKKGTGYLLVY-----LDHNDK 784
 QY 1176 LTIIVSAISNLIRKHVSEALVEDIGHELTLYVLYPAAREGAFVELLFEID----- 1226
 Db 785 RKMAY--LITDVCYHVEAEARGMHGOQIEILIPARKKE--FVLPQALEIQRNRYR 841
 QY 1227 -----DRLSDIGTSVSGISETTELEIFLKYAESGVDAETSDGTLPARNRRAV 1275
 Db 842 NVFDMNENTLKSQLATLEMRSTGLSINTLEQYFTITGDK-----VKAIAISRNSRIS 894
 QY 1276 GDSQSCARPTTEDDADPNDSIDIPESRETDLLSGMDGSGYVQVGMKLTQOQFVALLMK 1335
 Db 895 HNSRNASEPLKAGYDTQSTKSADSYOK--LMDQARKG--EKSQVAKMAQFISTMR 951
 QY 1336 RLIIARRSRKGFPAQIVLPAVFCIALVPSLIVPPFGKYPSLEDPMMYNYTFVSNDA 1395
 Db 952 KFLYSRRNMAQLETFQLIP--IILGLVNSL-----TLKSNNT 988
 QY 1396 PEDTGLLEILNALTKDPGCTGCMENPIPTPCQAGEEWTAAPVPTIMDLQANGMT 1455
 Db 989 DQ-----FSVRSI-----TP--SGIE-----PSKYVWRENGT-- 1014
 QY 1456 MONSPACQSSDKIKKMLPVCPCAGGLPPQKRONADLLODLTGRNISDYLVKTYVO 1515
 Db 1015 -----IPEAANEKILRRSGGEVLTNTKMLP 1044
 QY 1516 IIAKSLKAKKIWNERYGSGFLSVNTQALPPSOEVNDAIKMKHKLAKDSSADRFN 1575
 Db 1045 NTKSL-----IGE-----MPPA----- 1057
 QY 1576 SLGRFMTGLDTRNNKVMENKGMHAISFLNYNNALILANLQKGNPSHYITAENHP 1635
 Db 1058 TIGMTMNS-----DIEALFNRRYTHVLEPTLSMTNRRLGTDAETSSGVFLYSKSTN 1113
 QY 1636 LMLTQQLSEVALMTTSYDVLSICVIFAMSEVPASVFLIOEYRSKAKHLOFTLGVP 1695
 Db 1114 SNLPSQL-----IDVLLAMLLIFAMYSTFVMLIEBRCQARHQOFLTGISP 1164
 QY 1696 VTYMLSNFVMQNVVYPATLVITIFECPOQKSVSTNLPLVALLLLYGMSITPLMP 1755
 Db 1165 ITFFSASLIYGCITSLICILFLFELFAF---HMYDHLATVILFWLFFSPSVETIYA 1220
 QY 1756 ASEVFKIPSTAYVLTSVNLFTIGINSVATFVLELTPDKNLKINDILKSVL-----IF 1810
 Db 1221 VSFLEQSPASAVNLLIMQVVISGALLAVFLFM-----INIDEMLSILVINIFMELL 1275
 QY 1811 PHFLIGGLIDMKVQAMALRGENRVPSP---LSMDLVGRNLAMAVEGVFLI 1866
 Db 1276 PSYAFKGAII-----TINTYG---MILPSEELMMHDCGNAMLMGTFVGSFAL 1322
 QY 1867 TVLIQYRFFIR-----PRPVNAKLSPLNDE---DEVYRERORILLDGGQNDILEIK 1915
 Db 1323 FVLLOEFKVFARFLISQVATVBRSSHNNVQPMMDLPVCEVSESRERVRHVNSQNSALVI 1382
 QY 1916 ELTKIYRRKKRPAVDICVIGIPGCBFGLLVNGAGKSTFKMLTGDTYTRGDAFLKN 1975
 Db 1383 DLTKTR--GFTAVNELCLAVDQKCEFGLLGVNAGAKTTFFNILLGQSFASSGEMIGR 1440
 QY 1976 SILSNITHEVQNNGYCPQPDATITELLTGREHVEFPALLRGVPEKEVGVGWMALRKGLV 2035
 Db 1441 DVTELL-----SIGCPQPDALMDLITGRRESLEILLQMHGFENYKA--KALILCEVGM 1493

QY 1561 HLKADSSADRFNLGCFMTGLDTRNNKVMNNKGHAISSFLVNIINNALIRANLQK 1620
 Db 1121 -----TIGTITMS-----DNLELFNMYRYHLPPLTISMTNARLTGTVA 1161
 QY 1621 GBNPSHYGTATRNHPLNTLTKQOLSEVALMTTSVDVLSICVIFAMSEFPAFVFLIOER 1680
 Db 1162 EISSGVFLSKSTNSNLLPSOL-----IDVLLAPMLILIFAMVTSFFVFLIEER 1212
 QY 1681 VSKAKHLOFISGVKPVYIWLNSFWDMCNVVPATLVIIIFICQOQSYSTLPVLA 1740
 Db 1213 TCOFNAHQOGLTISPTTFYASLITDGLYSLICTILFPLAF-----HMYDHLATVIL 1268
 QY 1741 LLLYKWSITPLMYPAFVFKIPSTAVVLTSVNLFGINGSVAFVLEFPTDKLNIN 1800
 Db 1269 FEFLEFSSVPFIYAVSFLEQSPSKANVLLITVOVYISGALLAVFLIFM-----INID 1323
 QY 1801 DIKSVFL-----IPPHCLGKGLDMKKNAMDALERGENFVSP-----LSWDLVGR 1851
 Db 1324 EMLKSLIVNIIMFLLPSTAFSGAIT-----TINTYG-----MILPSEELMNMWHQCK 1370
 QY 1852 NLFAMAVEGVVEFLLTVDLQYRFFIR-----PRPVNAKSLPNDL-----DEDVRRRQ 1900
 Db 1371 NAMLMGTGVCSEFALFVLLQFKFVRRLSQYVTVRRSHNNVQPMKDDLPVCEVSSEER 1430
 QY 1901 RLDOGGONDLEIKELTKIYRKRKPAVDRIQVIRPGECFGLGVNAGKSSFTFKMLT 1960
 Db 1431 RHRVNSONSALVIRKDLTKFE--GRFAVNEELCLAVDQKRCFGLGVNAGKTTTNNLT 1488
 QY 1961 GDTTVTRDADFLNKNLSINIEVHOMGCPQPDATTELLTGREHVEFALLRGVPEKE 2020
 Db 1489 GOSFASSEAMIGRDVTELL-----SIGCPQPDALMDLTRESLEILIAQMHGENYK 1543
 QY 2021 VGKVGEMAIRKGLGVKGEKYGAGYSGNRRKLTAMALIGRPVVEFDEPTGMDPKAR 2080
 Db 1544 A--KAELLIECVGMIAHADLVRFVSGGOKRKISVGALLAPQMLIDEPTAGIDPKAR 1601
 QY 2081 RFLNMCALSVYKEGRS--VLTLSHMEPCFALCTRMALIMNGRCRIGSYQHLKNRPDGY 2139
 Db 1602 REVVELLIMCREHNSKALMTLSHMDCEALCSRIALVNRGSLIALIGSSOELKSLYGNV 1661
 QY 2140 TIVVRINGSNPDLKPVODFGLAFPGSVLKEKRRNM--LOYOLPSSLSL--ARLFSTLS 2195
 Db 1662 TMLSLTEPQORDMYVO--LVQTRLPNSVLKTTSTKTLMLKMOIPREKEDCSAFENVO 1720
 QY 2196 OSKRRLHIEDYVSQTTLDQVFNFA--KDQSDDLHLKDL 2234
 Db 1721 ALAKDLGVKDFPIIAQSSLETFURLAGLDLDDLDHSTVEIS 1762

RESULT 13

T00826
 hypobacterial protein T3266.22 - Arabidopsis thaliana (fragment)
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 12-Feb-1999 #sequence: revision 12-Feb-1999 #text: change 23-Mar-2001
 C:Accession: T00826
 R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaut
 submitted to the EMBL Data Library, November 1997
 A:Description: Arabidopsis thaliana chromosome II BAC T3266 genomic sequence.
 A:Reference number: 214163
 A:Accession: T00826
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1246 <R0U>
 A:Cross-References: EMBL:AC002510; NID:92618683; P1D:92618705
 A:Experimental source: cultivar Columbia
 C:Genetics:
 A:Map position: 2
 A:Intons: 33/3; 95/2; 113/3; 137/3; 168/3; 361/3; 421/2; 432/3; 493/3; 521/3; 535/3; 56
 A:Note: T3266.22

Query Match 11.9%; Score 1402; DB 2; Length 1246;
 Best Local Similarity 28.7%; Prog. NO. 4.3e-82;
 Matches 415; Conservative 195; Mismatches 388; Indels 450; Gaps 43;

QY 1019 VGLPSSKLSKTSQLSQSGMORKLSALAFVGGSKVVLIDETAGVDPYRRGRTWELLKY 1078
 Db 1 VGL-SDKINTLVLRALSGCMKRLSGIALIGNSKVILLDEPTSGMDPYSMRLTMOILIKI 59
 QY 1079 RQGRITLSTPHHMDADVLGRIALISHGKLCYGSLSFLKNOLGRTGYTTLVKRUVSS 1138
 Db 60 KKGRIITLTHSHMBAEFLCDRIGIMANGSLKCCGSSIFLKHNGVGYTLTLK----- 113
 QY 1139 LSSCNSSSTSYLKLKEISVQSSSDAGLSDHSDTLTIDVSAISNLIRKHVEARLVE 1198
 Db 114 -----TSPY-----VSAAHYIHRHIPSATCVS 136
 QY 1199 DIGHELTVPYPAKAGAFVLEFHEIDRL-----SDP-GISSYGSFTLEFELPKVAF 1253
 Db 137 EYGNISFRLP--LASLPCEFNMRLEISCMKNSDSYRQISQYIGSVTLEEVLRVA- 193
 QY 1254 ESGVDAEISDGTLPARRNRARFGDKQGLRPETEDDAADPNDSIDIPESRETDLSGMDG 1313
 Db 194 GCNLDIE-----DKQ-----EDIFVSPDTKSSLYIGISNQ 223
 QY 1314 KGSYQVK-----GWLITQ-----QGFVALLM----- 1334
 Db 224 KSMQPKLLASCNDAQAVITTSYAKAFRLIVAAMVLLIGFISIQCCGCSIIERSMEFWRHC 283
 QY 1335 -----KLLIARSRKGFQAQIVLPAVFQIALVFSILVPPGKYPSLELOPMWYNT 1389
 Db 284 KALPIKARASACHRDKRTVAFOFIIPAVFLLEGLF-----LQK- 323
 QY 1390 FVSNDAPEDEGTLELLNALKKDGFGRCMEGNPRTPDYTCQAGEBWTTAPVQITNLF 1449
 Db 324 --HPDQSTILTATVFNPLISGKG-----GGGPIPFD-----LSVPAKVAQYI 366
 QY 1450 QNG-----NMTQNPSPACOCSSDKIKKMLPVCPPAGLPPQKONTADLQDLGR 1503
 Db 367 EGGWIOPLRNTSYKFPNP-----KALADAI--DAAGP 397
 QY 1504 NISDYLKVTYVOILASLKNKIWNFEFRYSGSLVSNTOALPQSQEVNDAIKOMKHLK 1563
 Db 398 TLQPTLS-----MSEFLMSSFD-----QSYOSSRE-----G 424
 QY 1564 LAKDSSADRFNLGCFMTGLDTRNNKVMNNKGHAISSFLVNIINNALIRANLQGEN 1623
 Db 425 LSHSDCNHPDGLG-----YVLLHNGTQOHGPIYINMRAILR--LATGN- 470
 QY 1624 PSHYGTATRNHPLNTK--QOLSEVALMTTSVDVLSVTCVIFAMSEFPAFVFLIOERVS 1682
 Db 471 -KNMTIOTRNHPLPPTKQIORIQRHDLDAFSAALIYNI-----AFSFIAPSAFVPIVREKREV 525
 QY 1683 KAKHLOFISGVKPVYIWLNSFWDMCNVVPATLVIIIFICQOQSYSTLPVLA 1742
 Db 526 KAKHQQILISGVSVLSTYWDPISTFLPSTRAILLIFVAFGLEOFTIGIRPLPVLM 585
 QY 1743 LLYGWSITPLMPASVFKIPSTAVVLTSVNLFI-----GINGS 1782
 Db 586 LKYGALIASSTYCLIPFTFHSMAQATSSYSVLLPISLFWSESSNVLWVHFFSGILM 645
 QY 1783 VAFVLELF--VDNKLNNINID-----LKSVELIFPHFCGLKIDM--VKNOAMADALE 1833
 Db 646 VLSFVWGLIPATASANSYLKLELLFRYALQNFRLSPGFESDGLASIALIRGMDKSS 705
 QY 1834 RGENREVSPLSWDLVGNLFAMAVEGVVEFLLTVDLQYRFFIRPREVNAKLS----- 1886
 Db 706 H-----GVFEWNYTASICYLGLVLRLEYCRSMILTSF--HGIDIKSLIYTIQA 755
 QY 1887 -----PL-----NDEDEDVRRBORIILDGQONDILEIKE 1916
 Db 756 SLELELIDRYVSYSTFTEPLKSDGAIISDMEDIDVQDEBRDVAISGLSDTMFLYIGN 815
 QY 1917 LTKIYRRKR-----RPVADRICVIRPGECFGLGVNAGKSSFFKMLTGDYTRDADL 1972
 Db 816 LRKVPDGMKHNGRVAVAQSLFVSQAGECGFGLTGNAGKTTTLMISGEPTTSGTAFI 875

QY 1973 NKSILSNIEHONMGXCPQPDATITELLTGREHEVEFALLRGVPEKGVGEWAIRKL 2032
 Db 876 FGKDIYASKARQHGICGPOPDALFEYITVEHELEIARIKGVDDHNDVNEKLEEF 935
 QY 2033 GLVYKGEYAGNYSGGNKKRISTAMALLIGPPVFLDEPTGGDKARPLWN--CALSV 2090
 Db 936 DLKSHKSPFLSGGNKKRISVALIAMIGDPPILVDEPSTGMDPAKRFMDVIRSLST 995
 QY 2091 VEGGSVYITSHSMCECALCTRMALMVNRRCISGVQHKNRG----- 2136
 Db 996 RSGKTAIVLTTHSMNEQAQLCTRIGIMGRLICGSPQHLKTRGHNLELEVPYNGVK 1055
 QY 2137 -----DGYT-----I 2141
 Db 1056 PNEVSNVELENCOITQOMLFVNPQPRSLDGLDEVICGVSDSITPRDASASEISLSPEN 1115
 QY 2142 VVRIA-----GSPN-----DLKPV 2155
 Db 1116 VORIAKFLGNEORVSTLVPPLPEEDVRFDDQLEOLFRRGGIPLPILFAEMWLTKEKFSAL 1175
 QY 2156 QDFEGLAFPGSVLKEKRRMLQYOLPSSLS--SLARISILSOKKRILHEDYSQOTTL 2213
 Db 1176 DSTOSSPFGAIFKSCNGLSIKYQLPFGEGGLSLADAFGLHRRNRMLGIAEYSISOSTL 1235
 QY 2214 DOYFNFA 2221
 Db 1236 ETIFNFA 1243

RESULT 14
 T27121
 hypothetical protein Y53C10A.9 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
 C:Accession: T27121
 R:White, S.
 submitted to the EMBL Data Library, November 1998
 A:Reference number: 220314
 A:Accession: T27121
 A:status: preliminary; translated from GB/EMBL/DBJ
 A:molecule type: DNA
 A:Residues: 1-1564 <WIL>
 A:Cross-references: EMBL:AL033536; PIDN:CAA22142.1; CESP:Y53C10A.9
 A:Experimental source: clone Y53C10A
 C:Genetics:
 A:Gene: CESP:Y53C10A.9
 A:introns: 43/3; 92/2; 226/3; 354/1; 712/3; 817/1; 875/1; 916/3; 984/3; 1069/2; 1
 C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

Query Match 9.8%; Score 1158.5; DB 2; Length 1564;
 Best local Similarity 23.0%; Pred. No. 4, 2e-66;
 Matches 433; Conservative 324; Mismatches 673; Indels 451; Gaps 60;

QY 458 WTADT-----VAFLA-----KHPEVSSNGSV--YTREAFNET-----NO 493
 Db 22 WTGEVAIPILILIAFPAAMIVAKSOAPEDAKSSGSDIMMLSTESFHPTVSSTCDPSY 81
 QY 494 AIRTISFECVNLKLEPIATEVWLINKSMELDERKFAIGVFITPGSTELPHNV- 552
 Db 82 GVOITVNVENVYKMYSPKR-----W--DNTTNSLDLAKNPAKRLFTKDVACAKKIYIYE 133
 QY 553 -KYIRMDIDNVERTNKIKDGDWDPGRADPFEDMRVWVGFAV-----LQDVVEGA 603
 Db 134 DKYKTTWVLPAL-----GRRSYNFRSYNYNMKIMFGLSQRANNTSILENF 180
 QY 604 IIRVLGTE-KRTGVYMQMPYCYDDIFLRWMSRSMPLFMILAMTISAVAIKGIYE 662
 Db 181 DIRKYSTINEESG-----FGQNFGLFAVCMIMPIYSARALVYE 221
 QY 663 KEARLKETMRIMGLDNLFWFSFISLIPPLVSAGLLVILKLGNLLEPVSQSVFVFL 722
 Db 222 KSS-VKPYLTITGLP---LMMFIEHFLPGVINKTFLITLSTLYIFSMDNCTYVLGI 277

QY 723 SVFAVVTILLOCFLLISTLFSRANLAACGGIITFTLYLPVLCVA-----MDVY-----GF 773
 Db 278 FMYCHCVSILCTSLILPGKRIVEGMILIMTLLIAMLSTLEFEDMLFVPLNPV 337
 QY 774 TLKIFASLSEVAFGFCCEFAIFEEQIGVQMDNLFESVEEDFNLTSISMALPDF 833
 Db 338 SLKLFVATPLASPNQPTPSALFSSK-----KKTLOSAAVYEGIMISCTVMLVAALF 391
 QY 834 LYGVMTWIEAVFPFGQYGPWPWFPCTKSYW--FGESDEKS--HPCSNOKRMEICM 888
 Db 392 MEKLYTVGHAIF-----KRPWRLIGSKSKRKRKIERGCGVEDRSTILO 436
 QY 889 EEPETHLK---LGVSIONLVKYVDGMKYAVDGLALNFEGQITSLGNAGAKTTMS 944
 Db 437 CKETVEGSGALADIELSLGVKYVQNGEK-AVNGLSRAIRGOVSLILGNCGKSTFG 495
 QY 945 ILTGIFPPTSGTAYILKQDIRSEMTTRONLGVCPQNNVLFDMLTVEHEHWFARLGLS 1004
 Db 496 MITGMHQATEGKVMIGGIDANANRAEARIELGYCPQNPYDELTEWHELRVLNALKGRS 555
 QY 1005 -EKHYKAEMQMALDVGIPSSKLSKTSOLSGQMRKLSVALAFVGGSKVYILDEPTAGV 1063
 Db 556 GGSDFKMAESLKOIEL-TDKRNTLAKNLISGKRRKLCYCMAMIGSRVYLIDEPFAGM 614
 QY 1064 DPYSKRGIMELLKYRQGRITLITSTHMDADVLDGRIATISHGKLCVSSSLFLKNQIG 1123
 Db 615 DPARIDVQNNMALVYKADRTLLTHYDEAEKLGDMIFVMSHGKMAASGKHYLKQYG 674
 QY 1124 TGYITLV-----KRYESSLSGRSSSTVSYLKKEDBSOSSSDAGISDHS 1173
 Db 675 GGMILTIVFKSVHDPMPRKRYETAYDCKTVCSTA-LVKDER-----GQMIET 722
 QY 1174 DTLTIDVSAISNLRKRVSEARLVEDIGHETLYVLYEAAKEGAFVELHEIDRLSDLG 1233
 Db 723 SILTEKSRLLPTLK--LLESVMEEDYNNP-----EQALPEDIQKCTLE 767
 QY 1234 ISSGISSETTEIEIFLKYAEE-----SGVDAETSDGTLPARNRRAFGDKOSCLRPPT 1286
 Db 768 LATIGVSMSSLEOVFIKIGDECDIDMNGTGDKRT-----EROKKFTLV 812
 QY 1287 EDDADAPNDSDIDPESRRTDLISGMDKGSYOVKGWKLTOQOFALLMKRLLIARRSRK 1346
 Db 813 QYKIOQPK-----OGFSKLMMVVALLOKRAYVYLYRNPVQ 847
 QY 1347 FFAQIVLPAVEVCIALVSLIVPFGKYPSLELOPMWYNEQVTSNDAPEDTLELN 1406
 Db 848 ITLQILP-----LTLTILFAVPLR--LEPKPKRSLIESPDPQYHSTVLLQDEN 898
 QY 1407 ALTKDPGFGTRCMENGNPLPDTPCOAGEEMTTAPVPIIMDLFONGNWTMONPSPACCS 1466
 Db 899 E-----N 900
 QY 1467 SDIKKKMLPYCPGAGGLPPQROKNTADILQDLTGRNISDYLVKTYOIIAKSLKKIW 1526
 Db 901 DDLRLANLY-----NSFSNF-----EYVFKL----- 921
 QY 1527 VNEFRYGFSLGVSNTQALPSEOVNDAIKOMKKHLKAK--DSSADRFNLISGRFM-- 1561
 Db 922 -----GFIVYVKN-----KQDSKRYKISQGDKNALIMLIMTIASAMYLRD 960
 QY 1582 ---TGLDTRNNKVKVFNKGMHAISFLVNLINALLRANLQGENPSHYGTARNRHPLNL 1638
 Db 961 PSVTKLPVHVSVMIMND-----PRIKYEGIASF----- 989
 QY 1639 TKQQLSEVALMTTSVDVLSICVIFAMSFVPAVEVFLIOERVSARKHLOFISGVKPVY 1698
 Db 990 -----ELEENIFPLVLAGITQS-----TVYIEEKICKFAHDOYLTGISTAY 1034
 QY 1699 WLSNFWDMCNVYVATVITIFICEQOKSYVSTNLP--VLALLLLDY--WSITPLMY 1754
 Db 1035 WGVVFLMDE-----LLETFELLTYTIGFLISFVGLGHIHEIVIRYGLLFVAPLVY 1086
 QY 1755 PASFPFKIPSTAVVLTSVNLFIGINGSVATFVLELFDKNKLNINIDLKSVPLIF--PHF 1813

```

Db 1087 LNSALNTEPTRGNELL---YMFCCIPWLAAYSTVSELDHNEPPIÖKYSDEIEYGRIFENPSI 1143
QY 1814 CIGRGILDMVKNOAMADALEREENREVPSPISWDLVGRNLFAMAVEGVFFL----- 1865
Db 1144 GFLAGLIMKTAALNPKSGIDKHFEHL-----TNLM--TYEGIFEFELMFLFFGCI 1190
QY 1866 -ITVLI-----QYRFFI-----RPRYNKLSPLNDEDEYRREORRIIDGGGQNDIL 1912
Db 1191 FTILIGCATLKPFRNACRGTRRRSÖPERRRYKGISCAVKEEDLVQEVDNKNETVL 1250
QY 1913 EIKELTKIYRRKRPAYDRICVGPGEQGLLYGNAGKSSTEKMLTGDITVTRGDAFL 1972
Db 1251 VIDGLVKGDCGFR--AVNDLSISVGHCECFGLGANGAGKTTTIDITGLTPTGGSATI 1308
QY 1973 NKNSTLSNHEVHQMGCPOPDATTELLTGREHVEFPALLRGVPEKVGKVEWAIKRL 2032
Db 1309 DGHDTETELH----IGYCPQPDAMLQOISCRÖTLRIMAKLÖGYF--NVKEVELVLDV 1361
QY 2033 GLVYGEKYAGNYSGGNKRKLTAMALIGSPVYVLEDEPTGMDPKARRFLMNCALSYK 2092
Db 1362 GMSDFGYKLKNCSSGCKRKISVGIALMSRATCIIIDEPITAGIDPRARRIEMDIHEKRE 1421
QY 2093 EGR-SVLTSHSMEECALCTRMALMVNGRRFCIGSVÖHLKNRFGDGYTIIVRIAGSNPD 2151
Db 1422 QAKCSIVLTSHECEALCTRIGILRKEMIALGTSÖLSQYGNMYMFL-ILNSLED 1480
QY 2152 LKPVODFFGLAFPGSVLKEKHKRM---LÖYÖLPSLS--SLARIFSILSÖSKKRLHIEDYS 2207
Db 1481 LESVCVIVSEKMPDAVLTPSSSLTTSIVWELPKRSKDSKSEKYNÖVEVLAKKANAKDYM 1540
QY 2208 VSÖTTLDOVFVNFPAKQSDDD 2228
Db 1541 LTQASLEDFTFRLITTEDEEE 1561

```

RESULT 15

T47150

hypothetical protein DKFZp547P193.1 - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 02-Sep-2000

C:Accession: T47150

R:Bioecker, H.; Bioecker, M.; Brandt, P.; Mewes, H.W.; Well, B.; Wiemann, S.

submitted to the Protein Sequence Database, March 2000

A:Reference number: 224376

A:Accession: T47150

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-373 <AAA>

A:Cross-references: EMBL:AL162060

A:Experimental source: fetal brain; clone DKFZp547P193

A:Note: DKFZp547P193.1

C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

Query Match 8.68; Score 1013.5; DB 2; Length 373;

Best Local Similarity 65.0%; Pred. No. 8.4e-58;

Matches 195; Conservative 40; Mismatches 64; Indels 1; Gaps 1;

```

QY 1928 AVDRICVGPGEQGLLYGNAGKSSTEKMLTGDITVTRGDAFLMKNSTLSNHEVHON 1987
Db 7 AVDRICVGPGEQGLLYGNAGKSSTEKMLTGDITVTRGDAFLMKNSTLSNHEVHON 1987
QY 1988 MGYPQPDATTELLTGREHVEFPALLRGVPEKVGKVEWAIKRLGLVYGEKYAGNSG 2047
Db 67 LGYCPQPDALDELTAAREHQLYTRLRGISMKEARVYKMALEKLELTAKADPAAGTYSG 126
QY 2048 GNRKRLSTAMALIGSPVYVLEDEPTGMDPKARRFLMNCALSYVKEGSRVYLTSHSMEEC 2107
Db 127 GNRKRLSTAMALIGSPVYVLEDEPTGMDPKARRFLMNCALSYVKEGSRVYLTSHSMEEC 186
QY 2108 EALCTRLAIVNGRRLCLGSIQHLKNRFGDGYMIIVR-TRSSQSVKDVYRFFNRNPEAM 2167

```

```

Db 187 EALCTRLAIVNGRRLCLGSIQHLKNRFGDGYMIIVR-TRSSQSVKDVYRFFNRNPEAM 245
QY 2168 LKEKRRNMLÖYÖLPSLSLARIFSILSÖSKKRLHIEDYSVÖTTLDOVFVNFPAKQSDDD 2227
Db 246 LKEKRRNMLÖYÖLPSLSLARIFSILSÖSKKRLHIEDYSVÖTTLDOVFVNFPAKQSDDD 305

```

Search completed: March 7, 2003, 08:51:22
Job time: 121 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 7, 2003, 08:40:59 ; Search time 86 Seconds
(without alignments)
3503.252 Million cell updates/sec

Title: US-09-595-526c-2
Perfect score: 11797
Sequence: 1 MACMPEQLRLMKNLTFRRR.....VDVAVLTSFLQDEKVESYV 2261

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: A_Geneseq_101002:*

1: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:*

2: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:*

3: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:*

4: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:*

5: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:*

6: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:*

7: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:*

8: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:*

9: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:*

10: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:*

11: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:*

12: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:*

13: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:*

14: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:*

15: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:*

16: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:*

17: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:*

18: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:*

19: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:*

20: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:*

21: /SID52/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:*

22: /SID52/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:*

23: /SID52/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	11797	100.0	2261	22 AAB31361	Amino acid sequence
2	11797	100.0	2261	22 AAB31365	Amino acid sequence
3	11793	100.0	2261	21 AAB38117	Human ABC1 cholest
4	11793	100.0	2261	22 AAB31362	Amino acid sequence
5	11793	100.0	2261	22 AAB31366	Amino acid sequence
6	11792	100.0	2261	21 AAB38109	Human ABC1 cholest
7	11789	99.9	2261	21 AAB38082	Human ABC1 cholest
8	11789	99.9	2261	22 AAB71749	Human ABC1 protein
9	11789	99.9	2261	22 AAB31363	Amino acid sequence
10	11789	99.9	2261	22 AAB31367	Amino acid sequence

11	11786	99.9	2261	21 AAB38111	Human ABC1 cholest
12	11786	99.9	2261	21 AAB38114	Human ABC1 cholest
13	11786	99.9	2261	21 AAB38115	Human ABC1 cholest
14	11785	99.9	2261	21 AAB38105	Human ABC1 cholest
15	11785	99.9	2261	21 AAB38110	Human ABC1 cholest
16	11784	99.9	2261	21 AAB38113	Human ABC1 cholest
17	11784	99.9	2261	21 AAB38116	Human ABC1 cholest
18	11783	99.9	2261	21 AAB38112	Human ABC1 cholest
19	11777	99.8	2261	21 AAB38104	Human ABC1 cholest
20	11774.5	99.8	2261	21 AAB38106	Human ABC1 cholest
21	11771	99.8	2261	22 AAU02183	Human ABC1 mutant
22	11767	99.7	2259	21 AAB38107	Human ABC1 FFA-3 m
23	11767	99.7	2261	22 AAU02188	Human ABC1 mutant
24	11767	99.7	2261	22 AAU02189	Human ABC1 mutant
25	11767	99.7	2261	22 AAU02187	Human ABC1 mutant
26	11767	99.7	2261	22 AAU02186	Human ABC1 mutant
27	11767	99.7	2261	22 AAU02185	Human ABC1 mutant
28	11764	99.7	2261	22 AAU02184	Human ABC1 mutant
29	11763	99.7	2261	22 AAU02183	Human ABC1 mutant
30	11761	99.7	2261	22 AAU02182	Human ABC1 mutant
31	11760	99.7	2261	22 AAU02181	Human ABC1 mutant
32	11760	99.7	2261	22 AAU02180	Human ABC1 mutant
33	11760	99.7	2261	22 AAU02179	Human ABC1 mutant
34	11759	99.7	2261	22 AAU02178	Human ABC1 mutant
35	11759	99.7	2261	22 AAU02177	Human ABC1 mutant
36	11737	99.5	2261	23 AAB81578	Human ABC1 mutant
37	11433	96.9	2201	21 AAU02186	Human ABC1 mutant
38	11433	96.9	2201	21 AAU02185	Human ABC1 mutant
39	11433	96.9	2201	21 AAU02184	Human ABC1 mutant
40	11433	96.9	2201	21 AAU02183	Human ABC1 mutant
41	11015	93.4	2130	22 AAU02182	Human ABC1 mutant
42	7988	67.7	1525	22 AAU02181	Human ABC1 mutant
43	5858.5	49.6	1144	22 AAU02180	Human ABC1 mutant
44	5849	49.6	1144	22 AAU02179	Human ABC1 mutant
45	5849	49.6	1144	22 AAU02185	Human ABC1 mutant

ALIGNMENTS

RESULT 1

AA31361 standard; Protein: 2261 AA.

AC AAB31361:

20-APR-2001 (first entry)

DE Amino acid sequence of a human ABC1 polypeptide.

XX Human; adenosine triphosphate binding cassette protein 1; ABC1;

KW apolipoprotein-mediated mobilization; cholesterol; Tangier disease;

KW chromosome 9q22-q31; heart disease; hypercholesterolemia;

KW atherosclerosis; cholesterol transport.

OS Homo sapiens.

XX WO200078972-A2.

PN 28-DEC-2000.

XX 16-JUN-2000; 2000WO-US16765.

PF 18-JUN-1999; 99US-0140264.

PR 14-SEP-1999; 99US-0135872.

PR 19-NOV-1999; 99US-016573.

XX (CVTH-) CV THERAPEUTICS INC.

XX Lann RM, Wade D, Garvin M;

DR WPI; 2001-137812/14.

XX

PT Adenosine triphosphate (ATP) binding cassette (ABC) polynucleotide,
PT useful for the development of agents for the treatment of heart disease
PT and other disorders associated with hypercholesterolemia and
PT atherosclerosis -

PS Disclosure: Page 128-143; 215pp; English.

XX
CC The present sequence represents a human adenosine triphosphate (ATP)
CC binding cassette protein (ABC) 1 polypeptide. ABC1 resides in cell
CC membranes and utilizes ATP hydrolysis to transport a wide variety of
CC substrates across the plasma membrane. ABC1 is a pivotal protein in
CC the apolipoprotein-mediated mobilisation of intracellular cholesterol
CC stores. ABC1 is defective in Tangier disease, a genetic disorder
CC characterised by abnormal HDL-cholesterol metabolism. The ABC1 gene is
CC located on chromosome 9q22-9q31. The ABC1 genes and proteins are
CC useful for developing pharmaceutical agents for the treatment of heart
CC disease and other disorders associated with hypercholesterolemia and
CC atherosclerosis. The genes are useful for developing screening assays to
CC screen for compounds that regulate the expression of genes associated
CC with cholesterol transport. The genes and proteins are also useful for
CC are also useful as diagnostic indicators of cardiovascular disease and
CC other disorders associated with hypercholesterolemia.

XX Sequence 2261 AA:

Query Match 100.0%; Score 11797; DB 22; Length 2261;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MACMFOLELLMKNTFRRTOTQLELVAMPLEFILLISVLSLSPYRBOHCAPNKA 60
DB 1 MACMFOLELLMKNTFRRTOTQLELVAMPLEFILLISVLSLSPYRBOHCAPNKA 60
QY 61 MPSAGTLPVVOGIIICNANPCFRPTPGAPGVGNFNKSTIVARLPSDARLLIYSOKPT 120
DB 61 MPSAGTLPVVOGIIICNANPCFRPTPGAPGVGNFNKSTIVARLPSDARLLIYSOKPT 120
QY 121 SMKMRVYLFRTLOOIKSSSNLKLDPFLVNEFFSGFLYHNLSLPRSTYDKMLRADVILH 180
DB 121 SMKMRVYLFRTLOOIKSSSNLKLDPFLVNEFFSGFLYHNLSLPRSTYDKMLRADVILH 180
QY 181 KVFIOGVLHLTSLCNGSKSEMIOLGDOEVSFLCGLPEKLAERVLRSMDILKPL 240
DB 181 KVFIOGVLHLTSLCNGSKSEMIOLGDOEVSFLCGLPEKLAERVLRSMDILKPL 240
QY 241 RTLNSTSPFSKELAEATYKTLHSLGTLAQELFSSMSKSDMRQEVFLTNVSSSSSTQI 300
DB 241 RTLNSTSPFSKELAEATYKTLHSLGTLAQELFSSMSKSDMRQEVFLTNVSSSSSTQI 300
QY 301 YQAVSRIVCGHPGGGLIKSLNMYEDNNYKALFGGNGTEEDAEFFYNSTTPYCNOLMK 360
DB 301 YQAVSRIVCGHPGGGLIKSLNMYEDNNYKALFGGNGTEEDAEFFYNSTTPYCNOLMK 360
QY 361 NESSPSLRITMKALKELLVGLILYTPPTPATROYMAEVNKTFOELAVFHDLEGMWELIS 420
DB 361 NESSPSLRITMKALKELLVGLILYTPPTPATROYMAEVNKTFOELAVFHDLEGMWELIS 420
QY 421 PKIWTMENSOEMLVRLMLDSRDNDHFEOQLDGLMTADIDYAFIAKHEEDVOSSNGS 480
DB 421 PKIWTMENSOEMLVRLMLDSRDNDHFEOQLDGLMTADIDYAFIAKHEEDVOSSNGS 480
QY 481 YTWRAEFNEIQAIRITSRMECVNLKLEPIATEVWLINKSMELLDERKFMAGIYFTG 540
DB 481 YTWRAEFNEIQAIRITSRMECVNLKLEPIATEVWLINKSMELLDERKFMAGIYFTG 540
QY 541 ITPGSILEPHNHYKIRMDINVERTNKIKGYMDPGRADPPEDMRYVWGGEAYLDDV 600
DB 541 ITPGSILEPHNHYKIRMDINVERTNKIKGYMDPGRADPPEDMRYVWGGEAYLDDV 600
QY 601 ECAITIRVLTEKKTGYVMOQMPYCYVDDLEFLRVMSRSMPLMTLAWISVAVIITKGIY 660
DB 601 ECAITIRVLTEKKTGYVMOQMPYCYVDDLEFLRVMSRSMPLMTLAWISVAVIITKGIY 660

QY 661 YKEEARLKEFTMRIMGLDMSILMFSEISSLLPVLVSAGLLVILKGNLPPYSDPSVAFV 720
DB 661 YKEEARLKEFTMRIMGLDMSILMFSEISSLLPVLVSAGLLVILKGNLPPYSDPSVAFV 720
QY 721 FLTSVAVVTLIOCFLLISTFLSESRANLAACGGIIYFTYLPVLVLCVAAQDVGFILKIFMS 780
DB 721 FLTSVAVVTLIOCFLLISTFLSESRANLAACGGIIYFTYLPVLVLCVAAQDVGFILKIFMS 780
QY 781 ILSPAFPGGEYFALFEEOGIGVQMDNLESPVEEDGFNLJTISIMMLPDTFLYGVMTW 840
DB 781 ILSPAFPGGEYFALFEEOGIGVQMDNLESPVEEDGFNLJTISIMMLPDTFLYGVMTW 840
QY 841 YTEAVFPGOYGIIPRWYFPCRSYVFEESDEKSHPSNCKRMSFLCMEEPHILKLVGS 900
DB 841 YTEAVFPGOYGIIPRWYFPCRSYVFEESDEKSHPSNCKRMSFLCMEEPHILKLVGS 900
QY 901 IONLVKVRDGMKAVAVGGLALNFYEGOITSFLGHNGAGKTTMSILTGLPPTSGTAYTL 960
DB 901 IONLVKVRDGMKAVAVGGLALNFYEGOITSFLGHNGAGKTTMSILTGLPPTSGTAYTL 960
QY 961 GKDIRSEMSITRQMLGVCPQHNVLFDMLTYEHHIMFYARLKGISEKHVAEEMQNALDVG 1020
DB 961 GKDIRSEMSITRQMLGVCPQHNVLFDMLTYEHHIMFYARLKGISEKHVAEEMQNALDVG 1020
QY 1021 LPSSKLKSKTSQSLSGMQRRLSVALAFVSGSKVILDEPTAGADVPYSRGIWELLKTKRQ 1080
DB 1021 LPSSKLKSKTSQSLSGMQRRLSVALAFVSGSKVILDEPTAGADVPYSRGIWELLKTKRQ 1080
QY 1081 GRTIILSTHNMDEADVGLDRIATILSHGKLCVGSSTFLKNOGTGYTLTKKKDVESLS 1140
DB 1081 GRTIILSTHNMDEADVGLDRIATILSHGKLCVGSSTFLKNOGTGYTLTKKKDVESLS 1140
QY 1141 SCRRSSSTVSYLKKEDSVSSSSDAGLGSDBHSDTLTDVSAISMLIRKHSEARLVADI 1200
DB 1141 SCRRSSSTVSYLKKEDSVSSSSDAGLGSDBHSDTLTDVSAISMLIRKHSEARLVADI 1200
QY 1201 GHETIYVLPPEAKKEGAFELHEIDDRISDGLISYGISETTLEIFLKYAESGVDAE 1260
DB 1201 GHETIYVLPPEAKKEGAFELHEIDDRISDGLISYGISETTLEIFLKYAESGVDAE 1260
QY 1261 TSQCTLPARRNRRAFQKOSCLRPTEDDAADPNOSIDIPESRETDILSGMDGGSYOVK 1320
DB 1261 TSQCTLPARRNRRAFQKOSCLRPTEDDAADPNOSIDIPESRETDILSGMDGGSYOVK 1320
QY 1321 GAKTLQOQVALLMKRLILARRSRKGFRAQIYLPVAVYICALVPSLVPFGKXPSLEQ 1380
DB 1321 GAKTLQOQVALLMKRLILARRSRKGFRAQIYLPVAVYICALVPSLVPFGKXPSLEQ 1380
QY 1381 PMWYNQYTFVSNDAPEDTGTELELINALTKDPGCTRCMEGNPDPDPCOAGEEWTTAP 1440
DB 1381 PMWYNQYTFVSNDAPEDTGTELELINALTKDPGCTRCMEGNPDPDPCOAGEEWTTAP 1440
QY 1441 VQOTIMDLFQONNMTMOMPSPACOCSSDKIKKMLPVCBPAGGLPPROKRONADILLQDL 1500
DB 1441 VQOTIMDLFQONNMTMOMPSPACOCSSDKIKKMLPVCBPAGGLPPROKRONADILLQDL 1500
QY 1501 TGRNISDYLVKYYVOIILKSLKNTIWNVEFRYGGFSIGVSTQALPPRSQVNDATIKOMK 1560
DB 1501 TGRNISDYLVKYYVOIILKSLKNTIWNVEFRYGGFSIGVSTQALPPRSQVNDATIKOMK 1560
QY 1561 HLKALKSSADBFILNSLGRFMTGLDTRNNKWFNNKGMHAISFLAVINNALIRANLQK 1620
DB 1561 HLKALKSSADBFILNSLGRFMTGLDTRNNKWFNNKGMHAISFLAVINNALIRANLQK 1620
QY 1621 GENPSHYITAFNPNLNTKQOLSEVALMTTSVDVLSICVITFAMSFPASVFFLIQER 1680
DB 1621 GENPSHYITAFNPNLNTKQOLSEVALMTTSVDVLSICVITFAMSFPASVFFLIQER 1680
QY 1681 VSKAKHLQFISGVKPVITYLWLNFWDMCNYVPATVLIIFICFOOKSVYSSNTLPVIAL 1740
DB 1681 VSKAKHLQFISGVKPVITYLWLNFWDMCNYVPATVLIIFICFOOKSVYSSNTLPVIAL 1740
QY 1741 LLLLYGMSITPFLMYPASFVKIEPSTAVVLTISVNLFTGINSVAFFLEFLDNKLNNIN 1800

```

DB 1741 LLLLGWSTITPLMYPASAEVFKIPSTAYVLTISVNLFIGINSVAFFVLELTDKNLNNIN 1800
QY 1801 DLKSVFLIFEPFCGRLIDVKNQAMADALERGENRFPVSPISMDLVGRNLFMAAEG 1860
DB 1801 DLKSVFLIFEPFCGRLIDVKNQAMADALERGENRFPVSPISMDLVGRNLFMAAEG 1860
QY 1861 VVEFLITVLIQYRFRIRPPVNAKLSPLNDEDEDVREKORILDDGGONDILEIKELTKI 1920
DB 1861 VVEFLITVLIQYRFRIRPPVNAKLSPLNDEDEDVREKORILDDGGONDILEIKELTKI 1920
QY 1921 YRRKRPAVDRIQYRFRIRPPVNAKLSPLNDEDEDVREKORILDDGGONDILEIKELTKI 1980
DB 1921 YRRKRPAVDRIQYRFRIRPPVNAKLSPLNDEDEDVREKORILDDGGONDILEIKELTKI 1980
QY 1981 IHEVHONMGYCFQFADITELLTGREHVEFFALLRGVPEKEVGKGMIRKGLVKYCEK 2040
DB 1981 IHEVHONMGYCFQFADITELLTGREHVEFFALLRGVPEKEVGKGMIRKGLVKYCEK 2040
QY 2041 YAGNYSGNKRKLTAMALIGGPVVFLEDEPTTGMDPKARFLMNCALSVKEGRSVLT 2100
DB 2041 YAGNYSGNKRKLTAMALIGGPVVFLEDEPTTGMDPKARFLMNCALSVKEGRSVLT 2100
QY 2101 SHSMECEALCTRMATVNGRRCGLSVOLKKNRFGDGTIYVRIAGSNPDLKPVODFRG 2160
DB 2101 SHSMECEALCTRMATVNGRRCGLSVOLKKNRFGDGTIYVRIAGSNPDLKPVODFRG 2160
QY 2161 LAPPGVTLKRRHNMLOYOLPSSLSLARIFSTLSQSKRLHEDYSVSOTTLDOYFVNF 2220
DB 2161 LAPPGVTLKRRHNMLOYOLPSSLSLARIFSTLSQSKRLHEDYSVSOTTLDOYFVNF 2220
QY 2221 AKDQSDDDLKDLKSLHKNQTVVDVAVLTSFLQDEKYESYV 2261
DB 2221 AKDQSDDDLKDLKSLHKNQTVVDVAVLTSFLQDEKYESYV 2261

RESULT 2
AAB31365
ID AAB31365 standard; Protein: 2261 AA.
AC AAB31365;
XX 20-APR-2001 (first entry)
DE Amino acid sequence of a human ABC1 polypeptide.
XX Human: adenosine triphosphate binding cassette protein 1; ABC1;
KW apolipoprotein-mediated mobilisation; cholesterol; Tangier disease;
KW chromosome 9q22-9q31; heart disease; hypercholesterolemia;
XX atherosclerosis; cholesterol transport.
XX Homo sapiens.
XX WO200078971-A2.
XX 28-DEC-2000.
XX 16-JUN-2000; 2000WO-US16591.
XX 18-JUN-1999; 99US-0140264.
XX 14-SEP-1999; 99US-0153872.
XX 19-NOV-1999; 99US-0166573.
XX (CVTH-) CV THERAPEUTICS INC.
XX (UNIW) UNIT WASHINGTON.
XX Lawn RM, Wade D, Oram JF, Garvin M;
XX WPI: 2001-137811/14.
XX N-PSDB: AAF24702.
XX Adenosine triphosphate (ATP) binding cassette protein (ABC) 1
XX polynucleotides and polypeptides, useful for treatment of heart disease
PT

```

```

PT and other disorders associated with hypercholesterolemia and
PT atherosclerosis -
XX Claim 1: Page 123-138; 21pp; English.
XX The present sequence represents a human adenosine triphosphate (ATP)
CC binding cassette protein (ABC) 1 polypeptide. ABC1 resides in cell
CC membranes and utilizes ATP hydrolysis to transport a wide variety of
CC substrates across the plasma membrane. ABC1 is a pivotal protein in
CC the apolipoprotein-mediated mobilisation of intracellular cholesterol
CC stores. ABC1 is defective in Tangier disease, a genetic disorder
CC characterised by abnormal HDL-cholesterol metabolism. The ABC1 gene is
CC localised to chromosome 9q22-9q31. The ABC1 genes and proteins are
CC useful for developing pharmaceutical agents for the treatment of heart
CC disease and other disorders associated with hypercholesterolemia and
CC atherosclerosis. The genes are useful for developing screening assays to
CC screen for compounds that regulate the expression of genes associated
CC with cholesterol transport. The genes and proteins are also useful for
CC are also useful as diagnostic indicators of cardiovascular disease and
CC other disorders associated with hypercholesterolemia.
XX
XX Sequence 2261 AA:
SQ
Query Match 100.0%; Score 11797; DB 22; Length 2261;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MACWPQLRLMLKMLTFRROTCQLLEVPAMPLFILISVRLSYRPEDEHCFRKA 60
DB 1 MACWPQLRLMLKMLTFRROTCQLLEVPAMPLFILISVRLSYRPEDEHCFRKA 60
QY 61 MPSAGTLPEWVQITICNANPCFRYPTPEAGVYCNFNKSIVARLFSPARLLLYSQKD 120
DB 61 MPSAGTLPEWVQITICNANPCFRYPTPEAGVYCNFNKSIVARLFSPARLLLYSQKD 120
QY 121 SMKDMRKVLTLDQITKSSSNLKLQDFLVNDEFSGFLYHNLSPKSTVDKMLRADYILH 180
DB 121 SMKDMRKVLTLDQITKSSSNLKLQDFLVNDEFSGFLYHNLSPKSTVDKMLRADYILH 180
QY 181 KYFLOGYQLHLTSLCNGSKSEEMTQIGDOEVSELCGLPREKLAARVLRSMIDLKPL 240
DB 181 KYFLOGYQLHLTSLCNGSKSEEMTQIGDOEVSELCGLPREKLAARVLRSMIDLKPL 240
QY 241 RPLNSTSPPSKELAEATYTLTSLGTLAQLFMSRSMQDMQEWPLTNVSSSSSTOI 300
DB 241 RPLNSTSPPSKELAEATYTLTSLGTLAQLFMSRSMQDMQEWPLTNVSSSSSTOI 300
QY 301 YQAVSRITVCGHPGGGLIKSLNMYEDNNYKALFPGNGTEDEAFETFYDNTTPYCNDLMK 360
DB 301 YQAVSRITVCGHPGGGLIKSLNMYEDNNYKALFPGNGTEDEAFETFYDNTTPYCNDLMK 360
QY 361 NLESSPLSRITMKALKPLLVGKILYTPDTPATRQVMAEYNAKTFQELAVFHDEGWEELS 420
DB 361 NLESSPLSRITMKALKPLLVGKILYTPDTPATRQVMAEYNAKTFQELAVFHDEGWEELS 420
QY 421 PKITWMENSQENDLVBMILDSRDNDHEQDLDGDMTAODIYVFLAKHPEDVOSSNGS 480
DB 421 PKITWMENSQENDLVBMILDSRDNDHEQDLDGDMTAODIYVFLAKHPEDVOSSNGS 480
QY 481 VYTWREAFNETNOAIRTISFMECVMLNKLEPATEVWLINSMELDERKFMAGIVFTG 540
DB 481 VYTWREAFNETNOAIRTISFMECVMLNKLEPATEVWLINSMELDERKFMAGIVFTG 540
QY 541 ITPGSIELPHVKKYKLRMDIDNVERTNKIKDGYWDFGPRADEFEDRMRYWGGFAYLQDYV 600
DB 541 ITPGSIELPHVKKYKLRMDIDNVERTNKIKDGYWDFGPRADEFEDRMRYWGGFAYLQDYV 600
QY 601 EQALIRVLTGTEKKTGYVQMOQMPYPCYVDIPLRWRSRMPPLMTLAMYISAVAIITKGIY 660
DB 601 EQALIRVLTGTEKKTGYVQMOQMPYPCYVDIPLRWRSRMPPLMTLAMYISAVAIITKGIY 660
QY 661 YEKARLKETMRIMGDINSILMEFSITSLIPLVSAAGLVVLKGNLPLVSDPSVFEV 720
DB 661 YEKARLKETMRIMGDINSILMEFSITSLIPLVSAAGLVVLKGNLPLVSDPSVFEV 720

```

```

Db 661 YEKEARLKETMRIMGLDINSILMFESWFISSILPILVLVSAGLLVITLKGNLLEPSPSVFV 720
QY 721 FLSVFAVYVILLOCELLISTLFSRANI AAAAGGIIYFTLYLPVLCVAMODYVGTLLKIFAS 780
Db 721 FLSVFAVYVILLOCELLISTLFSRANI AAAAGGIIYFTLYLPVLCVAMODYVGTLLKIFAS 780
QY 781 LLSVPAFGGCEFFALFEEOGIGVOMDLFESPVBEDESFNTTISIMMLDPTLYGMV 840
Db 781 LLSVPAFGGCEFFALFEEOGIGVOMDLFESPVBEDESFNTTISIMMLDPTLYGMV 840
QY 841 YLEAVFPQGYGIPRPWYPPCTKSYWFESEDESKSHPGSNOKRMSJICMEEPHLLKGV 900
Db 841 YLEAVFPQGYGIPRPWYPPCTKSYWFESEDESKSHPGSNOKRMSJICMEEPHLLKGV 900
QY 901 IONLVKVRDGMKAVDGLALNFEYEGQTSFLGNAGKTTMSITGLPPTSGTAYIL 960
Db 901 IONLVKVRDGMKAVDGLALNFEYEGQTSFLGNAGKTTMSITGLPPTSGTAYIL 960
QY 961 GKDISEMSTIRONLGYCPOHNVLEFDMLTVEEHLMFYARLKGISEKHVKAEMQMALDV 1020
Db 961 GKDISEMSTIRONLGYCPOHNVLEFDMLTVEEHLMFYARLKGISEKHVKAEMQMALDV 1020
QY 1021 LPSSKLKSTKSOLSGGMQRKLSVALAVYGSKVYILDEPTAGVDPYSRRCIMWELLKRY 1080
Db 1021 LPSSKLKSTKSOLSGGMQRKLSVALAVYGSKVYILDEPTAGVDPYSRRCIMWELLKRY 1080
QY 1081 GRTIILSTHMDADVIGDRIATISHGKLCVSGSSFLKNOIGTYVTLVKKDVSSLS 1140
Db 1081 GRTIILSTHMDADVIGDRIATISHGKLCVSGSSFLKNOIGTYVTLVKKDVSSLS 1140
QY 1141 SCRNSSTVSTLKKEDSVSSSDAGLSGSHESDITLIDVSAISNLRKHVSARLVEDI 1200
Db 1141 SCRNSSTVSTLKKEDSVSSSDAGLSGSHESDITLIDVSAISNLRKHVSARLVEDI 1200
QY 1201 GHELTLYLPYFAKEGAFVELFHEIDRLSDLGISSETTLEELFKVAEESGVDAE 1260
Db 1201 GHELTLYLPYFAKEGAFVELFHEIDRLSDLGISSETTLEELFKVAEESGVDAE 1260
QY 1261 TSDGTLPARRRRAFGDKQSCLRPTEDDAADPNDSDIPESRETDLLSGMGKGSYQVK 1320
Db 1261 TSDGTLPARRRRAFGDKQSCLRPTEDDAADPNDSDIPESRETDLLSGMGKGSYQVK 1320
QY 1321 GSKTLTQOQFVALLMKRLIARRSRKGFPAQIVLPAVFCIALVFSCLVPPCKKYSLEQ 1380
Db 1321 GSKTLTQOQFVALLMKRLIARRSRKGFPAQIVLPAVFCIALVFSCLVPPCKKYSLEQ 1380
QY 1381 PMMYNEQYTFVNDAPEDTGTLELNAITKDPGFGTRCEGPNIPDTPQOAGEEEMTAP 1440
Db 1381 PMMYNEQYTFVNDAPEDTGTLELNAITKDPGFGTRCEGPNIPDTPQOAGEEEMTAP 1440
QY 1441 VPOTIMDLFONGWMTQONPSPAQCQSSDKIKMLPVCPPGAGGLPBPORKONTADILQDL 1500
Db 1441 VPOTIMDLFONGWMTQONPSPAQCQSSDKIKMLPVCPPGAGGLPBPORKONTADILQDL 1500
QY 1501 TGRNISDYLVKTYVOJIAKSLKNK IWNNEFRYGGFSLGYSNQALPPOEVNDAIKOMKK 1560
Db 1501 TGRNISDYLVKTYVOJIAKSLKNK IWNNEFRYGGFSLGYSNQALPPOEVNDAIKOMKK 1560
QY 1561 HUKLADSSADRFNLGFRPMGLDTRNNVYKWFNNKGHAHSFLNVIINNAILRANLQK 1620
Db 1561 HUKLADSSADRFNLGFRPMGLDTRNNVYKWFNNKGHAHSFLNVIINNAILRANLQK 1620
QY 1621 GENPSHYGITAENHPLNTKQOLSEVALMTTSVDVLVSCIYFAMSFPASVFFLQER 1680
Db 1621 GENPSHYGITAENHPLNTKQOLSEVALMTTSVDVLVSCIYFAMSFPASVFFLQER 1680
QY 1681 VSKAKHLOEISGVKPYIYVLSNFWDMCNVYVPATLVIIIFCPOOKSYASTNLPVAL 1740
Db 1681 VSKAKHLOEISGVKPYIYVLSNFWDMCNVYVPATLVIIIFCPOOKSYASTNLPVAL 1740
QY 1741 LLLLYGWSITPLMTFASVYFKIPSTAYVYVTSVNLFIGINGSVATFVLELFTDKNLNN 1800
Db 1741 LLLLYGWSITPLMTFASVYFKIPSTAYVYVTSVNLFIGINGSVATFVLELFTDKNLNN 1800

```

```

QY 1801 DILKSVPLIPPHFCGLRGLIDMVKNOAMADALEFFGENRVSPLSMDLVGRNLFAMA VEG 1860
Db 1801 DILKSVPLIPPHFCGLRGLIDMVKNOAMADALEFFGENRVSPLSMDLVGRNLFAMA VEG 1860
QY 1861 VFFELITVLIOYREFIRPRPVNAKLSPUNDEDEVRERORILIDGGONDILEIKELTKI 1920
Db 1861 VFFELITVLIOYREFIRPRPVNAKLSPUNDEDEVRERORILIDGGONDILEIKELTKI 1920
QY 1921 YRRKRKPADVDRICVGIIPRGECEGLGYNGAGKSTFEMLTGDTTVTRGDAFLKNKSI 1980
Db 1921 YRRKRKPADVDRICVGIIPRGECEGLGYNGAGKSTFEMLTGDTTVTRGDAFLKNKSI 1980
QY 1981 IHEVQNMNGYCPQPDATTELTLGREHEFFALLRGVPEKVGCVGMALRKILGVKYGK 2040
Db 1981 IHEVQNMNGYCPQPDATTELTLGREHEFFALLRGVPEKVGCVGMALRKILGVKYGK 2040
QY 2041 YAGNYSGNKRRKSTAMALIGGPPVFLDEPTTGMDPKARRFLMNCALSVEKGRSVLT 2100
Db 2041 YAGNYSGNKRRKSTAMALIGGPPVFLDEPTTGMDPKARRFLMNCALSVEKGRSVLT 2100
QY 2101 SHSMECEALCTRMAIWNNGRFICSGVOHLKNRFGDGTIVVRIAGSNPDIAKPYQDFEG 2160
Db 2101 SHSMECEALCTRMAIWNNGRFICSGVOHLKNRFGDGTIVVRIAGSNPDIAKPYQDFEG 2160
QY 2161 LAFPGSVLKEKRRNMLQYOLPSSLSIARIFSILSOSKKRLHIEDYSVQOTLIDQVNF 2220
Db 2161 LAFPGSVLKEKRRNMLQYOLPSSLSIARIFSILSOSKKRLHIEDYSVQOTLIDQVNF 2220
QY 2221 AKDOSDDHLKDLISLHKNQTVNVAVLTSFLQDEKESYV 2261
Db 2221 AKDOSDDHLKDLISLHKNQTVNVAVLTSFLQDEKESYV 2261

RESULT 3
AAB38117
ID AAB38117 standard; protein: 2261 AA.
XX
AC AAB38117;
XX
DT 29-JAN-2001 (first entry)
XX
DE Human ABC1 cholesterol transporter mutant, 1883M.
XX
KW Human ABC1 cholesterol transporter; chromosome 9q31;
KW ATP-binding cassette; HDL deficiency disorder; high density lipoprotein;
KW Tangier disease; TD; familial HDL deficiency; FHL; polymorphism;
KW cardiovascular disease; coronary artery disease; coronary restenosis;
KW cerebrovascular disease; peripheral vascular disease;
KW Alzheimer's disease; Niemann-Pick disease; Huntington's disease;
KW X-linked adrenoleukodystrophy; cancer; gene therapy; genetic diagnosis;
KW prognosis; prophylaxis; drug screening; transgenic animal; mutant;
KW muten.
XX
OS Homo sapiens.
XX
PN WO20005318-A2.
XX
PD 21-SEP-2000.
XX
PE 15-MAR-2000; 2000WO-1B00532.
XX
PR 15-MAR-1999; 99US-0124702.
PR 08-JUN-1999; 99US-0138048.
PR 17-JUN-1999; 99US-0139600.
PR 01-SEP-1999; 99US-0151977.
XX
PA (UVRB-) UNIV BRITISH COLUMBIA.
PA (XENO-) XENON BIORESEARCH INC.
XX
Hayden MR, Wilson AR, Pimstone SN;
WPI; 2000-587528/55.

```

XX New ABC1 polypeptide is useful for treating diseases associated with
 PT ABC1 biological activity, e.g. Alzheimer's disease, Huntington's
 PT disease and cancer -

PS Examples; Page - 1: 229pp; English.

XX The invention relates to the human ABC1 cholesterol transporter protein
 CC (B3082) and to nucleic acid sequences (C69120) which encode it. ABC1 is
 CC a member of the ATP-binding cassette (ABC transporter) superfamily of
 CC proteins, and plays a crucial role in cholesterol transport, particularly
 CC intracellular cholesterol trafficking in monocytes and fibroblasts, being
 CC involved in cholesterol efflux from the cell. The gene encoding ABC1 is
 CC located on chromosome 9q31, and mutations in this gene are associated
 CC with two genetic HDL (high density lipoprotein) deficiency disorders,
 CC Tangier disease (TD) and familial HDL deficiency (FHD). These diseases
 CC are distinguishable in that TD is an autosomal dominant trait, while
 CC FHD is inherited as an autosomal dominant trait. Low levels of HDL ("good
 CC cholesterol") in the blood correlate with a high risk of cardiovascular
 CC disease, particularly coronary artery disease, but also cerebrovascular
 CC disease, coronary restenosis, and peripheral vascular disease.
 CC Conversely, a high level of HDL has protective effects against
 CC cardiovascular disease. The invention provides genetic constructs and
 CC transgenic cells and non-human animals comprising human ABC1 nucleic
 CC acids, and methods of gene therapy for the treatment or prevention of
 CC cardiovascular disease comprising the administration of an expression
 CC vector encoding ABC1 or an active fragment thereof. The invention also
 CC encompasses compounds which mimic ABC1 activity, compounds which
 CC stimulate ABC1 expression and methods of screening for such compounds.
 CC It further relates to methods for determining whether a patient has an
 CC increased risk for cardiovascular disease due to polymorphisms in the
 CC ABC1 gene. Human ABC1 proteins and nucleotides can be used to treat
 CC or prevent cardiovascular disease, especially coronary artery disease,
 CC cerebrovascular disease, coronary restenosis or peripheral vascular
 CC disease. They may also be used in the treatment of diseases associated
 CC with ABC1 biological activity, such as Alzheimer's disease, Niemann-Pick
 CC disease, Huntington's disease, X-linked adrenoleukodystrophy and cancer.
 CC The invention specifically excludes proteins with the exact amino acid
 CC sequences of Genbank Accession No. CAH10005.1 and X75926.1. The
 CC acid with the exact sequence as Genbank Accession No. AJ012376.1. The
 CC present sequence represents a mutant human ABC1 cholesterol transporter
 CC associated with an altered cholesterol level and therefore an altered
 CC risk of cardiovascular disease.
 CC Note: The present sequence is not shown in the specification, but is
 CC derived from the native human ABC1 shown on pages 152-157.

CC Sequence 2261 AA:

Query Match 100.0%; Score 11793; DB 21; Length 2261;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 2259; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MACWPQLRLIMKMLTFRROTQCLLEVAMPFLFILILSVLSPVPEQHCHPNKA 60
 DB 1 MACWPQLRLIMKMLTFRROTQCLLEVAMPFLFILILSVLSPVPEQHCHPNKA 60
 QY 61 MPAGTLPMWOGIICNANPCFRPTPEGEAGVYGVNFKSIVARLSDARLLLYSQKPT 120
 DB 61 MPAGTLPMWOGIICNANPCFRPTPEGEAGVYGVNFKSIVARLSDARLLLYSQKPT 120
 QY 121 SMKDKRKYVLTLOQIKRSSNMLKQDELVDNETFSGLFYNNLSLPKSTYDKMKRADVILH 180
 DB 121 SMKDKRKYVLTLOQIKRSSNMLKQDELVDNETFSGLFYNNLSLPKSTYDKMKRADVILH 180
 QY 181 KYFLOGYGLHLTSLCNGSKSEEMTOLGDOEYSELGLPKREKLAEEVYLSNMDILKPTL 240
 DB 181 KYFLOGYGLHLTSLCNGSKSEEMTOLGDOEYSELGLPKREKLAEEVYLSNMDILKPTL 240
 QY 241 RTLNSTSPFSPKELAEAKTTLHSLGTLAQLFMSRMSMDRQFVMEFLTVNVSSTSTOT 300
 DB 241 RTLNSTSPFSPKELAEAKTTLHSLGTLAQLFMSRMSMDRQFVMEFLTVNVSSTSTOT 300
 QY 301 YQAVSRIVCGHPEGGGLIKSLNLTENNNYKALFGNGTDEDAETFYDNSTTPYCNDLKM 360

DB 301 YQAVSRIVCGHPEGGGLIKSLNLTENNNYKALFGNGTDEDAETFYDNSTTPYCNDLKM 360
 QY 361 NLESSPLSRITWAKMLPLVYKILYTPDPTRQVMAEVNKTFOELAVHDLBGMWEELS 420
 DB 361 NLESSPLSRITWAKMLPLVYKILYTPDPTRQVMAEVNKTFOELAVHDLBGMWEELS 420
 QY 421 PKIWTFMENSOBMLVRLNLDSDNDHFVHQQLDMDTADIVAFIAKHPEVQSSNGS 480
 DB 421 PKIWTFMENSOBMLVRLNLDSDNDHFVHQQLDMDTADIVAFIAKHPEVQSSNGS 480
 QY 481 VYTWAEAFNETMOAIRTSRPEMECNLKLPEIATEVWLIINKSMELLDERKFWAGIVFTG 540
 DB 481 VYTWAEAFNETMOAIRTSRPEMECNLKLPEIATEVWLIINKSMELLDERKFWAGIVFTG 540
 QY 541 ITPGSTLEPHVKKYIRMDIDNVEKTKIKGYNDPPGRADPFEDMRYVWGFAVLDQV 600
 DB 541 ITPGSTLEPHVKKYIRMDIDNVEKTKIKGYNDPPGRADPFEDMRYVWGFAVLDQV 600
 QY 601 EQAIIIRVLTGTEKTKGVYMOQMPYPCYVDIIFLRMSRSMPLFTLAMIYSVAVILKIV 660
 DB 601 EQAIIIRVLTGTEKTKGVYMOQMPYPCYVDIIFLRMSRSMPLFTLAMIYSVAVILKIV 660
 QY 661 YEKEARLKEFMRIKGLDNLIMFSWFISSLPLVYAGLLVILKGNLPLSDPSVVF 720
 DB 661 YEKEARLKEFMRIKGLDNLIMFSWFISSLPLVYAGLLVILKGNLPLSDPSVVF 720
 QY 721 FLISVFAVNTILQCFILSTLFSRANLAACGGIYFTLPLVLCVAMODYGFLTKIFAS 780
 DB 721 FLISVFAVNTILQCFILSTLFSRANLAACGGIYFTLPLVLCVAMODYGFLTKIFAS 780
 QY 781 LLSVFAFGCEYFALFEBOGIGVOMDLNLFESVPEDEFNLTTSIMMLPDTLYGVWTW 840
 DB 781 LLSVFAFGCEYFALFEBOGIGVOMDLNLFESVPEDEFNLTTSIMMLPDTLYGVWTW 840
 QY 841 YIEAVFPQYGIIPRWYFPTQTSYNGEESDKSIHPSNOKMSEICMEEPHILKIGVS 900
 DB 841 YIEAVFPQYGIIPRWYFPTQTSYNGEESDKSIHPSNOKMSEICMEEPHILKIGVS 900
 QY 901 IONLVKVRDGMVAADGALANFEGQITSPFLGHNAGKTTMSILTGLPPTSGTAVIL 960
 DB 901 IONLVKVRDGMVAADGALANFEGQITSPFLGHNAGKTTMSILTGLPPTSGTAVIL 960
 QY 961 GKDIRSEMSTIRONLGVCPQHNVLFMPLVBEHILYTAALKLSKHNKAEBOALDVG 1020
 DB 961 GKDIRSEMSTIRONLGVCPQHNVLFMPLVBEHILYTAALKLSKHNKAEBOALDVG 1020
 QY 1021 LPSSKLSKTSLSGSGMQRKLSVALAFVGSKVYILDEPTAGVDPSRKGIELLKYRQ 1080
 DB 1021 LPSSKLSKTSLSGSGMQRKLSVALAFVGSKVYILDEPTAGVDPSRKGIELLKYRQ 1080
 QY 1081 GRTIILSTHMHDEADVLDGRIAIISHGKLCVGSFLKNOLOCTGYTLTVKKDVESLS 1140
 DB 1081 GRTIILSTHMHDEADVLDGRIAIISHGKLCVGSFLKNOLOCTGYTLTVKKDVESLS 1140
 QY 1141 SCRNSSTVSTYTKKEDSVSSDAGISGDHESDPTLIDVSAISNLRKHVSEARLVEDI 1200
 DB 1141 SCRNSSTVSTYTKKEDSVSSDAGISGDHESDPTLIDVSAISNLRKHVSEARLVEDI 1200
 QY 1201 GHETLYVLPYPAKEGAFVELFHEIDRLSDLGISSYGISSETTLEELFLKVAEESGVDAE 1260
 DB 1201 GHETLYVLPYPAKEGAFVELFHEIDRLSDLGISSYGISSETTLEELFLKVAEESGVDAE 1260
 QY 1261 TSDGTLPAKRRNRAFGDKOSCLRPETEDDAADPDSDSIDESSETDLDSMDKSGSYOK 1320
 DB 1261 TSDGTLPAKRRNRAFGDKOSCLRPETEDDAADPDSDSIDESSETDLDSMDKSGSYOK 1320
 QY 1321 GSKLQOQFVALWKLRLIARSRKGFPAQIVLPAVVCIALVFSLLVPFGKYPSELD 1380
 DB 1321 GSKLQOQFVALWKLRLIARSRKGFPAQIVLPAVVCIALVFSLLVPFGKYPSELD 1380
 QY 1381 PMVYNQYTFVSNDAPEDTGTLLELNLTKDQFGTRCMQENPPTPCQAGEEWTPAP 1440
 DB 1381 PMVYNQYTFVSNDAPEDTGTLLELNLTKDQFGTRCMQENPPTPCQAGEEWTPAP 1440

QY 301 YQAVSRIYCGHEPGGGLIKISLNMVEDNNYKALFGNGTDEEDATFTYDNSTTPYCNIDLMK 360
 Db 301 YQAVSRIYCGHEPGGGLIKISLNMVEDNNYKALFGNGTDEEDATFTYDNSTTPYCNIDLMK 360
 QY 361 NNESSPLRIITKALKPLLVGKILYTPPTPATROYMAEVNKTFOELAVFHLEGMWELLS 420
 Db 361 NNESSPLRIITKALKPLLVGKILYTPPTPATROYMAEVNKTFOELAVFHLEGMWELLS 420
 QY 421 PKIWTFMENSOEMDLVRLMLDSRDNDHFEQOOLDJMTAODIYAFIAKHEBDYQSSNGS 480
 Db 421 PKIWTFMENSOEMDLVRLMLDSRDNDHFEQOOLDJMTAODIYAFIAKHEBDYQSSNGS 480
 QY 481 VYTWEAEAFNETQARTISREMECVNLKLPRIATEVWLVKSMELDERKFMAGIYPTG 540
 Db 481 VYTWEAEAFNETQARTISREMECVNLKLPRIATEVWLVKSMELDERKFMAGIYPTG 540
 QY 541 TTPGSELPHHVKYKLRMDIDNVERTNKIKDGYMDPCPRADPFEDMRYVMGCFAYLQDV 600
 Db 541 TTPGSELPHHVKYKLRMDIDNVERTNKIKDGYMDPCPRADPFEDMRYVMGCFAYLQDV 600
 QY 601 EQAIIIRVLGTGTEKKTGYVMQMPYCYVDJFLVAVSRSMPLMTLAMIYSAVVIKGIY 660
 Db 601 EQAIIIRVLGTGTEKKTGYVMQMPYCYVDJFLVAVSRSMPLMTLAMIYSAVVIKGIY 660
 QY 661 YKKEARKKTKMIRMGIDNSILMFESWFISSLPLVYASGLVVIKLGMLLPYSDPSVYEV 720
 Db 661 YKKEARKKTKMIRMGIDNSILMFESWFISSLPLVYASGLVVIKLGMLLPYSDPSVYEV 720
 QY 721 PLSEFAVVTIICQFLISTFSSRANLAACGGLIYFTLYLVYLCVAMODYVGTFTKIFAS 780
 Db 721 PLSEFAVVTIICQFLISTFSSRANLAACGGLIYFTLYLVYLCVAMODYVGTFTKIFAS 780
 QY 781 LLSVPAFGGCEYFALFEEQIGVQMDNI.FESPVEBDGFNLTTSSIMLFDFTLYGVMTW 840
 Db 781 LLSVPAFGGCEYFALFEEQIGVQMDNI.FESPVEBDGFNLTTSSIMLFDFTLYGVMTW 840
 QY 841 YTEAVPFGQYGI.PRPYFPCSTKSYMGEESDEKSHPGSNOKRMSFCMEDEEPTHLKLGVS 900
 Db 841 YTEAVPFGQYGI.PRPYFPCSTKSYMGEESDEKSHPGSNOKRMSFCMEDEEPTHLKLGVS 900
 QY 901 IONLVKVRDGMKAVADGALNFEQOITSPFGHNGAKTTMSILTGLPPTSGTAYIL 960
 Db 901 IONLVKVRDGMKAVADGALNFEQOITSPFGHNGAKTTMSILTGLPPTSGTAYIL 960
 QY 961 GKDIREBEMSTIRONLGVCGQHNVLPDMLTVEBEHIMFATLKGISEKHVAAEMQALDVG 1020
 Db 961 GKDIREBEMSTIRONLGVCGQHNVLPDMLTVEBEHIMFATLKGISEKHVAAEMQALDVG 1020
 QY 1021 LPSKLSKTSOLSGMOKRLSVALAFVGSKVYLDEPTAGVDPRSRGIMELLKTYRQ 1080
 Db 1021 LPSKLSKTSOLSGMOKRLSVALAFVGSKVYLDEPTAGVDPRSRGIMELLKTYRQ 1080
 QY 1081 GRTIILSTHHMDEADVLDGRIALISHGKLCVGGSSILFKNQLGSTGYTLVYKKNDESSLS 1140
 Db 1081 GRTIILSTHHMDEADVLDGRIALISHGKLCVGGSSILFKNQLGSTGYTLVYKKNDESSLS 1140
 QY 1141 SCNNSSTVYLLKKEEDSVQSSDAGLGDHESDTLITIVSAISNLIKRAHVSARLVEDI 1200
 Db 1141 SCNNSSTVYLLKKEEDSVQSSDAGLGDHESDTLITIVSAISNLIKRAHVSARLVEDI 1200
 QY 1201 GHBLTYLVLEPAKKEGAFVELFHEIDRLSDLGISSYGISETTLEELFKVAEESGVDAE 1260
 Db 1201 GHBLTYLVLEPAKKEGAFVELFHEIDRLSDLGISSYGISETTLEELFKVAEESGVDAE 1260
 QY 1261 TSDGTLPARNRNRAFSGDKSLRPFTTEDAADPNDSIDIPESSETFLVLSGMCKGKYQV 1320
 Db 1261 TSDGTLPARNRNRAFSGDKSLRPFTTEDAADPNDSIDIPESSETFLVLSGMCKGKYQV 1320
 QY 1321 GMLTQOOFVALLMKRLILARSRKGFPAQIYLPVAVFCIALVFSLIVPEPGKYPSELEIO 1380
 Db 1321 GMLTQOOFVALLMKRLILARSRKGFPAQIYLPVAVFCIALVFSLIVPEPGKYPSELEIO 1380

QY 1381 PMWNEQYTFVNSDADPEDTGLELLNALITKDPGFCRCMEGNIPDTPCOAGEEEMTAP 1440
 Db 1381 PMWNEQYTFVNSDADPEDTGLELLNALITKDPGFCRCMEGNIPDTPCOAGEEEMTAP 1440
 QY 1441 VPQTIMDLFQNGMWTQONPSPACQCSSDKIKKMLPVCPCPAGSLPPQKQONTADILQDL 1500
 Db 1441 VPQTIMDLFQNGMWTQONPSPACQCSSDKIKKMLPVCPCPAGSLPPQKQONTADILQDL 1500
 QY 1501 TGRNISDYLVKTYVOIITAKSLKNIWNEFRYCGFSLGYSNTQALPPSOEVNDAIKOMK 1560
 Db 1501 TGRNISDYLVKTYVOIITAKSLKNIWNEFRYCGFSLGYSNTQALPPSOEVNDAIKOMK 1560
 QY 1561 HLKIAKSSADREPLNSIGRPMGTGLDFRNNVKVWPNKGMHAISSPLNVLINNALIRANLQK 1620
 Db 1561 HLKIAKSSADREPLNSIGRPMGTGLDFRNNVKVWPNKGMHAISSPLNVLINNALIRANLQK 1620
 QY 1621 GENSHYGITAFNHPNLMTQOOLSEVALMTTSDVLSICVIFAMSFVPASFPVFLIOER 1680
 Db 1621 GENSHYGITAFNHPNLMTQOOLSEVALMTTSDVLSICVIFAMSFVPASFPVFLIOER 1680
 QY 1681 VSKAKHLOFTSGVPVITYWLSNFWQMCNIVVPATVITITFICFOOKSVYSTNLPLVAL 1740
 Db 1681 VSKAKHLOFTSGVPVITYWLSNFWQMCNIVVPATVITITFICFOOKSVYSTNLPLVAL 1740
 QY 1741 LLLLYGMSITPPLMYPASFEVKIPSTAVVVLTSVNLFGINGSVATEVLELTDNKLNIN 1800
 Db 1741 LLLLYGMSITPPLMYPASFEVKIPSTAVVVLTSVNLFGINGSVATEVLELTDNKLNIN 1800
 QY 1801 DILKSVFLIPHECLRGKIDMWKQAMADALBERFGENRFVSPSLMDVGRNLFMAAVEG 1860
 Db 1801 DILKSVFLIPHECLRGKIDMWKQAMADALBERFGENRFVSPSLMDVGRNLFMAAVEG 1860
 QY 1861 VVEFLITVLYOYREFIRPRPVNAKLSPLNDEDEVRERORIIIDGGGQNDILEIKELTKI 1920
 Db 1861 VVEFLITVLYOYREFIRPRPVNAKLSPLNDEDEVRERORIIIDGGGQNDILEIKELTKI 1920
 QY 1921 YRRKRKPAVDRIKIGVIRPGECFGLLGYNAGKSSFTKMLTGDTTVTREGDAFLKNKSTLSN 1980
 Db 1921 YRRKRKPAVDRIKIGVIRPGECFGLLGYNAGKSSFTKMLTGDTTVTREGDAFLKNKSTLSN 1980
 QY 1981 IHEVHQNMGYCPQDATTETLLTGREHVEFPALLRGVBEKEVKYGEWAIKKLGIVKXGK 2040
 Db 1981 IHEVHQNMGYCPQDATTETLLTGREHVEFPALLRGVBEKEVKYGEWAIKKLGIVKXGK 2040
 QY 2041 YAGNYSKGNRKRLSTAMALIGCPVYVFLDEPTTGMDCARARFLMNCALSVYKEGRSVLT 2100
 Db 2041 YAGNYSKGNRKRLSTAMALIGCPVYVFLDEPTTGMDCARARFLMNCALSVYKEGRSVLT 2100
 QY 2101 SHSMECEALCTRNAIMVNGRFCLGSVOHLKNRFGDGYTTIVARIAGSNPDLKPVQDFG 2160
 Db 2101 SHSMECEALCTRNAIMVNGRFCLGSVOHLKNRFGDGYTTIVARIAGSNPDLKPVQDFG 2160
 QY 2161 LAFPSVYLKEKHNRLQYOLPSSLSLARIFSILSOSKRRHIEDYSVSOTTLDOVYVF 2220
 Db 2161 LAFPSVYLKEKHNRLQYOLPSSLSLARIFSILSOSKRRHIEDYSVSOTTLDOVYVF 2220
 QY 2221 AKDQSDDDLKDLSLHKQTVADVAVLTSFLQDEKVESYV 2261
 Db 2221 AKDQSDDDLKDLSLHKQTVADVAVLTSFLQDEKVESYV 2261

RESULT 5
 AAB31366
 ID AAB31366 standard; Protein: 2261 AA.
 XX
 AC AAB31366:
 XX
 DT 20-Apr-2001 (first entry)
 XX
 DE Amino acid sequence of ABC1 polypeptide from Tangier disease patient.
 XX Human: adenosine triphosphate binding cassette protein 1; ABC1:
 KM apolipoprotein-mediated mobilisation; cholesterol; Tangier disease;

Qy	1321	GWKLTQOQFALLMKRLLLIARSRKCFEFOIYLPVAVFCIALVSLIYPRGKXPSLEQ	1380
Db	1321	GWKLTQOQFALLMKRLLLIARSRKCFEFOIYLPVAVFCIALVSLIYPRGKXPSLEQ	1380
Qy	1381	PMWNEOYTVSNDADBEDGTLELLNALTKDGFGRCMEGNPIDPTPCQAGEEMTTAP	1440
Db	1381	PMWNEOYTVSNDADBEDGTLELLNALTKDGFGRCMEGNPIDPTPCQAGEEMTTAP	1440
Qy	1441	VPQTMDLFONGWMTYQONSPPACQCSSDKIKKMLPYCPRGAGLEPPOROKONTADIIQDL	1500
Db	1441	VPQTMDLFONGWMTYQONSPPACQCSSDKIKKMLPYCPRGAGLEPPOROKONTADIIQDL	1500
Qy	1501	TGRNISDYLTQVYQIIAASLKKIWNNEFRYGGFSLGYSNTQALPPOEYNDAIKQMK	1560
Db	1501	TGRNISDYLTQVYQIIAASLKKIWNNEFRYGGFSLGYSNTQALPPOEYNDAIKQMK	1560
Qy	1561	HLKLAKDSADREFLNSLGREMTGLDTRNNKYWFNNKGMAHAISSFLVYINNAIIIRANLQK	1620
Db	1561	HLKLAKDSADREFLNSLGREMTGLDTRNNKYWFNNKGMAHAISSFLVYINNAIIIRANLQK	1620
Qy	1621	GENSHGCIATFANHPLNLTQKQQLSEVALMTTSDVLYSICVIFAMSPFASVYVFLQER	1680
Db	1621	GENSHGCIATFANHPLNLTQKQQLSEVALMTTSDVLYSICVIFAMSPFASVYVFLQER	1680
Qy	1681	VSRAKHLOFTSGVKPYLYWMLSNFWMQCNVVPATVLIITFICFOOKSVSSTNLPLAL	1740
Db	1681	VSRAKHLOFTSGVKPYLYWMLSNFWMQCNVVPATVLIITFICFOOKSVSSTNLPLAL	1740
Qy	1741	LLLLGMSITPLMTPASVFKITSTAYVLYTSVNLFTIGNSVATFVLELPFDNKLNIN	1800
Db	1741	LLLLGMSITPLMTPASVFKITSTAYVLYTSVNLFTIGNSVATFVLELPFDNKLNIN	1800
Qy	1801	DILKSVELIPPHFCGLGLDMYKNOAMADALREFCENRFVSPILSMDLYGRULFMAAEG	1860
Db	1801	DILKSVELIPPHFCGLGLDMYKNOAMADALREFCENRFVSPILSMDLYGRULFMAAEG	1860
Qy	1861	VVFPLITVLIQYREFIRPRPVNAKLSPLNDEDEVDREBORILLDGGQNDILEIKELTKI	1920
Db	1861	VVFPLITVLIQYREFIRPRPVNAKLSPLNDEDEVDREBORILLDGGQNDILEIKELTKI	1920
Qy	1921	YRRRRKPAVNDICIGIRPGCFGLGYNAGSKSTPKMLTGDYTVYRGDAFLNKNSILSN	1980
Db	1921	YRRRRKPAVNDICIGIRPGCFGLGYNAGSKSTPKMLTGDYTVYRGDAFLNKNSILSN	1980
Qy	1981	IHEYHONNGYCPQOFADITELLTGRREHVEFPALLIGVPEKEVGKVGEMAIIRKLGIVKYEK	2040
Db	1981	IHEYHONNGYCPQOFADITELLTGRREHVEFPALLIGVPEKEVGKVGEMAIIRKLGIVKYEK	2040
Qy	2041	YAGNYSGNKKRKLSTAAALLIGSPRYVLEDEPTTGMPKRRFLMNCALSIVEGKSVYLT	2100
Db	2041	YAGNYSGNKKRKLSTAAALLIGSPRYVLEDEPTTGMPKRRFLMNCALSIVEGKSVYLT	2100
Qy	2101	SHSMECECALOTRMAIIVNGRFRCLGSVQHLKRNFPDGYTIIVIRIAGSNPDLKPVQDFEG	2160
Db	2101	SHSMECECALOTRMAIIVNGRFRCLGSVQHLKRNFPDGYTIIVIRIAGSNPDLKPVQDFEG	2160
Qy	2161	LAFPGSVLKEKRRMLDYOQLPSSISLARIFSLSSKKRLHIEDYSVSQTTLDQVYVNF	2220
Db	2161	LAFPGSVLKEKRRMLDYOQLPSSISLARIFSLSSKKRLHIEDYSVSQTTLDQVYVNF	2220
Qy	2221	AKDOSDDPHLKDLSLHKNOYVNDVAVLTSFLQDEKKESEYV 2261	
Db	2221	AKDOSDDPHLKDLSLHKNOYVNDVAVLTSFLQDEKKESEYV 2261	
RESULT 6			
AAB38109			
ID	AAB38109 standard; Protein: 2261 AA.		
XX			
AC	AAB38109:		
XX			
DT	29-JAN-2001 (first entry)		

DE	Human ABC1 cholesterol transporter mutant, R219K.
XX	
KW	Human ABC1 cholesterol transporter: chromosome 9q31;
KM	ATP-binding cassette; HDL deficiency disorder; high density lipoprotein;
KW	Tangier disease; TD; familial HDL deficiency; FHA; polymorphism;
KM	cardiovascular disease; coronary artery disease; coronary restenosis;
KW	cerebrovascular disease; peripheral vascular disease;
KM	Alzheimer's disease; Niemann-Pick disease; Huntington's disease;
KW	X-linked adrenoleukodystrophy; cancer; gene therapy; genetic diagnosis;
KM	prognosis; prophylaxis; drug screening; transgenic animal; mutant;
XX	mutelin.
XX	
OS	Homo sapiens.
XX	
PN	W0200055318-A2.
PD	
XX	21-SEP-2000.
XX	
PF	15-MAR-2000; 2000WO-1B00532.
XX	
PR	15-MAR-1999; 99US-0124702.
PR	08-JUN-1999; 99US-0138048.
PR	17-JUN-1999; 99US-0139600.
PR	01-SEP-1999; 99US-0151977.
XX	
PA	(UYBR-) UNIV BRITISH COLUMBIA.
PA	(XENO-) XENON BIORESEARCH INC.
XX	
PI	Hayden MR, Wilson AR, Pimstone SN;
DR	WPI; 2000-587528/55.
XX	
PT	New ABC1 polypeptide is useful for treating diseases associated with
PT	ABC1 biological activity, e.g. Alzheimer's disease, Huntington's
PT	disease and cancer -
XX	
PS	Examples: Page -: 22ppp; English.
XX	
CC	The invention relates to the human ABC1 cholesterol transporter protein
CC	(B38082) and to nucleic acid sequences (C69120) which encode it. ABC1 is
CC	a member of the ATP-binding cassette (ABC transporter) superfamily of
CC	proteins, and plays a crucial role in cholesterol transport, particularly
CC	intracellular cholesterol trafficking in monocytes and fibroblasts, being
CC	involved in cholesterol efflux from the cell. The gene encoding ABC1 is
CC	located on chromosome 9q31, and mutations in this gene are associated
CC	with two genetic HDL (high density lipoprotein) deficiency disorders,
CC	Tangier disease (TD) and familial HDL deficiency (FHA). These diseases
CC	are distinguishable in that TD is an autosomal recessive disorder, while
CC	FHA is inherited as an autosomal dominant trait. Low levels of HDL ("good
CC	cholesterol") in the blood correlate with a high risk of cardiovascular
CC	disease, particularly coronary artery disease, but also cerebrovascular
CC	disease, coronary restenosis, and peripheral vascular disease.
CC	Conversely, a high level of HDL has protective effects against
CC	cardiovascular disease. The invention provides genetic constructs and
CC	transgenic cells and non-human animals comprising human ABC1 nucleic
CC	acids, and methods of gene therapy for the treatment or prevention of
CC	cardiovascular disease comprising the administration of an expression
CC	vector encoding ABC1 or an active fragment thereof. The invention also
CC	encompasses compounds which mimic ABC1 activity, compounds which
CC	stimulate ABC1 expression and methods of screening for such compounds.
CC	It further relates to methods for determining whether a patient has an
CC	increased risk for cardiovascular disease due to polymorphisms in the
CC	ABC1 gene. Human ABC1 proteins and nucleotides can be used to treat
CC	or prevent cardiovascular disease, especially coronary artery disease,
CC	cerebrovascular disease, coronary restenosis or peripheral vascular
CC	disease. They may also be used in the treatment of diseases associated
CC	with ABC1 biological activity, such as Alzheimer's disease, Niemann-Pick
CC	disease, Huntington's disease, X-linked adrenoleukodystrophy and cancer.
CC	The invention specifically excludes proteins with the exact amino acid
CC	sequences of GenBank accession No: CAA10005.1 and X75926, and the nucleic
CC	acid with the exact sequence as GenBank accession No: AJ012376.1. The
CC	present sequence represents a mutant human ABC1 cholesterol transporter
CC	associated with an altered cholesterol level and therefore an altered

CC Risk of cardiovascular disease.

CC Note: The present sequence is not shown in the specification, but is derived from the native human ABC1 shown on pages 152-157.

XX Sequence 2261 AA:

Query Match 100.0%; Score 11792; DB 21; Length 2261;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2259; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MACWPOLRLMLKNTLFRROTCCOLLLEVAMPLEFLITISVRLSYRPEYEOHECFPKA 60
DB 1 MACWPOLRLMLKNTLFRROTCCOLLLEVAMPLEFLITISVRLSYRPEYEOHECFPKA 60
QY 61 MSAGTLPMVVOGICNANPCFRYPYPCGAPGVGNPKSIVARLFSARILLYSQKDT 120
DB 61 MSAGTLPMVVOGICNANPCFRYPYPCGAPGVGNPKSIVARLFSARILLYSQKDT 120
QY 121 SKMKRKYVRLTLOQIKKSSSNLKLODPLVDNETFSGFLYHNLSTLPKSTVDKMLRADVILH 180
DB 121 SKMKRKYVRLTLOQIKKSSSNLKLODPLVDNETFSGFLYHNLSTLPKSTVDKMLRADVILH 180
QY 181 KVFLOQYOLHLTSLONGSKSEEMIQLDQEVSELGFLPKKELAAAEVRLSRMMDILKPL 240
DB 181 KVFLOQYOLHLTSLONGSKSEEMIQLDQEVSELGFLPKKELAAAEVRLSRMMDILKPL 240
QY 241 RLNSTSPRSKELAEARKTLLHSLGTLAQLFMSRMSDKRQEVMLTNNSSSSSTQI 300
DB 241 RLNSTSPRSKELAEARKTLLHSLGTLAQLFMSRMSDKRQEVMLTNNSSSSSTQI 300
QY 301 YQAVSRIVCGHPEGGLIKSLNWEEDNNYKALFGNGCEEDAEFTYNSITTPYCNDLML 360
DB 301 YQAVSRIVCGHPEGGLIKSLNWEEDNNYKALFGNGCEEDAEFTYNSITTPYCNDLML 360
QY 361 NLESSPLSRITIKALKPLLVGKILYPTDPATROYMAEVNTPQELAVHDLLEGMEELS 420
DB 361 NLESSPLSRITIKALKPLLVGKILYPTDPATROYMAEVNTPQELAVHDLLEGMEELS 420
QY 421 PKIWFEMESQEMDLVRMLDLSRDNHDEOOLGDMTADIVAFIAKHEDVOSSSGS 480
DB 421 PKIWFEMESQEMDLVRMLDLSRDNHDEOOLGDMTADIVAFIAKHEDVOSSSGS 480
QY 481 VYTWAEAFNETNOALRTISREMECVNLKLEPIATEVWLINKSMELDEREFMAIGVFTG 540
DB 481 VYTWAEAFNETNOALRTISREMECVNLKLEPIATEVWLINKSMELDEREFMAIGVFTG 540
QY 541 ITPGSELPHHYKXKIRMDINDVERTNKIKGYMDPGPRADPFEDMRYWVGFAVLQDVV 600
DB 541 ITPGSELPHHYKXKIRMDINDVERTNKIKGYMDPGPRADPFEDMRYWVGFAVLQDVV 600
QY 601 EQAIIIRVLGTEKKTGVYMOOMPYPYCYDDJFLRYMSRSMPLFMILAMTYSVAVILKGI 660
DB 601 EQAIIIRVLGTEKKTGVYMOOMPYPYCYDDJFLRYMSRSMPLFMILAMTYSVAVILKGI 660
QY 661 YEKARLKEETMRIMGLDINSILMFSSISLLPLVYAGLIVILKGLNLPYSDPVVFF 720
DB 661 YEKARLKEETMRIMGLDINSILMFSSISLLPLVYAGLIVILKGLNLPYSDPVVFF 720
QY 721 FLVSFAVVTIIIOCLISLFSRANLAAAGGIIYFTLLPVLCVANOVDYVGLKLPAS 780
DB 721 FLVSFAVVTIIIOCLISLFSRANLAAAGGIIYFTLLPVLCVANOVDYVGLKLPAS 780
QY 781 LLSPAVFGGCEYFALFEEOGIGVOMDNLFESEPEDEDFNLITTSISMLEFPTFLYGVMTW 840
DB 781 LLSPAVFGGCEYFALFEEOGIGVOMDNLFESEPEDEDFNLITTSISMLEFPTFLYGVMTW 840
QY 841 YIEAFVPGQYGIIPRWTFPCIKSTWFGESDEKSHPGSNOKRMSEICMEEPHTLKGVS 900
DB 841 YIEAFVPGQYGIIPRWTFPCIKSTWFGESDEKSHPGSNOKRMSEICMEEPHTLKGVS 900
QY 901 IONLVKVRDGMKYAVDGLALNFYEGQITSLFGHNGAKTMTSLILGLPPTSGTAYIL 960
DB 901 IONLVKVRDGMKYAVDGLALNFYEGQITSLFGHNGAKTMTSLILGLPPTSGTAYIL 960

QY 961 GKDIRSEMTIIRQNLGYCPOHNVLFDMIVVEEHIWFAARLKGSEKHVKAEMQMALDVG 1020
DB 961 GKDIRSEMTIIRQNLGYCPOHNVLFDMIVVEEHIWFAARLKGSEKHVKAEMQMALDVG 1020
QY 1021 LPSSKLKSKTSQLSGCKORLISVALAEVGGSKVYLIDEPTAGVDPYSRKGIMWELLKRYRO 1080
DB 1021 LPSSKLKSKTSQLSGCKORLISVALAEVGGSKVYLIDEPTAGVDPYSRKGIMWELLKRYRO 1080
QY 1081 GRTITLSTHMDADVGDGLIILSHGKILCCVSSSLFIKNQIGTGYLLTLKKDVSSIS 1140
DB 1081 GRTITLSTHMDADVGDGLIILSHGKILCCVSSSLFIKNQIGTGYLLTLKKDVSSIS 1140
QY 1141 SCRRSSSTVSYLKKEDSVSSQSSDAGLSGDSHEDTLTIDVSAISNLIRKHVSARLVEDI 1200
DB 1141 SCRRSSSTVSYLKKEDSVSSQSSDAGLSGDSHEDTLTIDVSAISNLIRKHVSARLVEDI 1200
QY 1201 GHETLTVLPYEAKEGAFVFLFHEIDRLSDLGISYGISSETTLEBFLKVAEESGVDAE 1260
DB 1201 GHETLTVLPYEAKEGAFVFLFHEIDRLSDLGISYGISSETTLEBFLKVAEESGVDAE 1260
QY 1261 TSDGTLPARNRRAFGDKQCLPFTEDDAADPNDSIDIPESRETDLLSGMGKGSYQK 1320
DB 1261 TSDGTLPARNRRAFGDKQCLPFTEDDAADPNDSIDIPESRETDLLSGMGKGSYQK 1320
QY 1321 GWKLTQOQFVALLMKRLILARRSRKGFPAQIYLPAAVFCIALVSLIYPPFKRYSLELO 1380
DB 1321 GWKLTQOQFVALLMKRLILARRSRKGFPAQIYLPAAVFCIALVSLIYPPFKRYSLELO 1380
QY 1381 PMWYNEGYFVSDAPEDDTLELNLALTKDPGFGRCMEGNIPPTPCQAGEEETTPAP 1440
DB 1381 PMWYNEGYFVSDAPEDDTLELNLALTKDPGFGRCMEGNIPPTPCQAGEEETTPAP 1440
QY 1441 VPQTIMDLFONGMWTQONSPPACQSSDKIKKMLPVCPPGAGGLPPOKQKMTADILQDL 1500
DB 1441 VPQTIMDLFONGMWTQONSPPACQSSDKIKKMLPVCPPGAGGLPPOKQKMTADILQDL 1500
QY 1501 TGRNISYLYKTYVOIITIASLKNKIVNBEFRYGGSLVSNQIALPPSOEVDAIKOMK 1560
DB 1501 TGRNISYLYKTYVOIITIASLKNKIVNBEFRYGGSLVSNQIALPPSOEVDAIKOMK 1560
QY 1561 HLKLAKSSADRFUNSLGRMTGLDTRNNVYKWFNNKGHAISFLNVNNAILKANLOK 1620
DB 1561 HLKLAKSSADRFUNSLGRMTGLDTRNNVYKWFNNKGHAISFLNVNNAILKANLOK 1620
QY 1621 GENPSHYGITAENHPLNLTQOOLSEVALMTTSVDVLSICVIFAMSEFVPSFVFLQER 1680
DB 1621 GENPSHYGITAENHPLNLTQOOLSEVALMTTSVDVLSICVIFAMSEFVPSFVFLQER 1680
QY 1681 VSKAKHLQFTSGYKPYIYMLSNFVWDMCNVVPATLVIIIFCFQOKSVSSSTNLPVAL 1740
DB 1681 VSKAKHLQFTSGYKPYIYMLSNFVWDMCNVVPATLVIIIFCFQOKSVSSSTNLPVAL 1740
QY 1741 LLLLYGMSITPLMYPASFEFKIPSTAYVVLTSVNLFIIGINGSVAFVLELFTDNKLNNIN 1800
DB 1741 LLLLYGMSITPLMYPASFEFKIPSTAYVVLTSVNLFIIGINGSVAFVLELFTDNKLNNIN 1800
QY 1801 DILKSVFLIFPHFCGLIDWYKKNQAMADALBERGENRFVSPLSMDLVGRNLFMAAVEG 1860
DB 1801 DILKSVFLIFPHFCGLIDWYKKNQAMADALBERGENRFVSPLSMDLVGRNLFMAAVEG 1860
QY 1861 VVFFLITVLIQYRFFIRPPVNAKLSPLNDEDEDDVRRERQRLIDGGQNDILEIKELTKI 1920
DB 1861 VVFFLITVLIQYRFFIRPPVNAKLSPLNDEDEDDVRRERQRLIDGGQNDILEIKELTKI 1920
QY 1921 YRRKRRPAVDRLICVGIIPGECGGLGVNGAGKSSFFKMLTGTDTYTRGDAFLNKSIIISN 1980
DB 1921 YRRKRRPAVDRLICVGIIPGECGGLGVNGAGKSSFFKMLTGTDTYTRGDAFLNKSIIISN 1980
QY 1981 IHEVHQMNGYQOPDAITELLTGREHVEFFALLRGVPEKEVGKVEBMAIRKLGIVKYEK 2040
DB 1981 IHEVHQMNGYQOPDAITELLTGREHVEFFALLRGVPEKEVGKVEBMAIRKLGIVKYEK 2040

QY 2041 YAGNYSGNKKRLSTAMALIGPPVVFLEDEPTGMDPKARFPLWNCALSVYKESRSVLT 2100
 DB 2041 YAGNYSGNKKRLSTAMALIGPPVVFLEDEPTGMDPKARFPLWNCALSVYKESRSVLT 2100
 QY 2101 SHSMECEALCTRMALWNGRFRGCGVQHLKKNRGGDCTTVIRIAGSNPDLKRVQDFPG 2160
 DB 2101 SHSMECEALCTRMALWNGRFRGCGVQHLKKNRGGDCTTVIRIAGSNPDLKRVQDFPG 2160
 QY 2161 LAFPSVLEKEKRRNMLQYOLPSSLSLARIFSLSQSKRLHIEDYSVSQTTLDQVEVNF 2220
 DB 2161 LAFPSVLEKEKRRNMLQYOLPSSLSLARIFSLSQSKRLHIEDYSVSQTTLDQVEVNF 2220
 QY 2221 AKDQSDDDLKDLSTLHKQTVYDAVLTSLQDEKVFESYV 2261
 DB 2221 AKDQSDDDLKDLSTLHKQTVYDAVLTSLQDEKVFESYV 2261
 RESULT 7
 AAB38082
 ID AAB38082 standard; Protein: 2261 AA.
 AC AAB38082;
 XX
 DT 29-JAN-2001 (first entry)
 XX
 DE Human ABC1 cholesterol transporter.
 XX
 XX Human ABC1 cholesterol transporter; chromosome 9q31;
 KW ATP-binding cassette; HDL deficiency disorder; high density lipoprotein;
 KW Tangier disease; TD; familial HDL deficiency; FHA; polymorphism;
 KW cardiovascular disease; coronary artery disease; coronary restenosis;
 KW cerebrovascular disease; peripheral vascular disease;
 KW Alzheimer's disease; Niemann-Pick disease; Huntington's disease;
 KW X-linked adrenoleukodystrophy; cancer; gene therapy; genetic diagnosis;
 KW prognosis; prophylaxis; drug screening; transgenic animal.
 XX
 OS Homo sapiens.
 XX
 PN MO200055318-A2.
 XX
 PD 21-SEP-2000.
 XX
 PF 15-MAR-2000; 2000WO-IB00532.
 XX
 PR 15-MAR-1999; 99US-0124702.
 PR 08-JUN-1999; 99US-0138048.
 PR 17-JUN-1999; 99US-0139600.
 PR 01-SEP-1999; 99US-0151977.
 XX
 PA (UYBR-) UNIV BRITISH COLUMBIA.
 PA (XENO-) XENON BIORESEARCH INC.
 XX
 PI Hayden MR, Wilson AR, Pimstone SN;
 XX
 DR WPI: 2000-587528/55.
 DR N-PSDB: AAC69120.
 XX
 PT New ABC1 polypeptide is useful for treating diseases associated with
 PT ABC1 biological activity, e.g. Alzheimer's disease, Huntington's
 PT disease and cancer.
 XX
 PS Claim 5; Page 152-157; 229pp; English.
 XX
 CC The invention relates to the human ABC1 cholesterol transporter protein
 CC (AB38082) and to nucleic acid sequences (C69120) which encode it. ABC1 is
 CC a member of the ATP-binding cassette (ABC transporter) superfamily of
 CC proteins, and plays a crucial role in cholesterol transport, particularly
 CC intracellular cholesterol trafficking in monocytes and fibroblasts, being
 CC involved in cholesterol efflux from the cell. The gene encoding ABC1 is
 CC located on chromosome 9q31, and mutations in this gene are associated
 CC with two genetic HDL (high density lipoprotein) deficiency disorders,
 CC Tangier disease (TD) and familial HDL deficiency (FHA). These diseases
 CC are distinguishable in that TD is an autosomal recessive disorder, while

CC FHA is inherited as an autosomal dominant trait. Low levels of HDL ("good
 CC cholesterol") in the blood correlate with a high risk of cardiovascular
 CC disease, particularly coronary artery disease, but also cerebrovascular
 CC disease, coronary restenosis, and peripheral vascular disease.
 CC Conversely, a high level of HDL has protective effects against
 CC cardiovascular disease. The invention provides genetic constructs and
 CC transgenic cells and non-human animals comprising human ABC1 nucleic
 CC acids, and methods of gene therapy for the treatment or prevention of
 CC cardiovascular disease comprising the administration of an expression
 CC vector encoding ABC1 or an active fragment thereof. The invention also
 CC encompasses compounds which mimic ABC1 activity, compounds which
 CC stimulate ABC1 expression and methods of screening for such compounds.
 CC It further relates to methods for determining whether a patient has an
 CC increased risk for cardiovascular disease due to polymorphisms in the
 CC ABC1 gene. Human ABC1 proteins and nucleotides can be used to treat
 CC or prevent cardiovascular disease, especially coronary artery disease,
 CC cerebrovascular disease, coronary restenosis or peripheral vascular
 CC disease. They may also be used in the treatment of diseases associated
 CC with ABC1 biological activity, such as Alzheimer's disease, Niemann-Pick
 CC disease, Huntington's disease, X-linked adrenoleukodystrophy and cancer.
 CC The invention specifically excludes proteins with the exact amino acid
 CC sequences of GenBank Accession No: CAA10005.1 and X75926, and the nucleic
 CC acid with the exact sequence as GenBank Accession No: AJ012376.1. The
 CC present sequence represents the human ABC1 cholesterol transporter.
 XX
 SQ Sequence 2261 AA:
 Query Match 99.9%; Score 11789; DB 21; Length 2261;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 2258; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MACMPQLRLMLKMLNTRFRROTCQLLEVAMPFLIFLILISRLSYRPPYDQHECHFNKA 60
 DB 1 MACMPQLRLMLKMLNTRFRROTCQLLEVAMPFLIFLILISRLSYRPPYDQHECHFNKA 60
 QY 61 MPSAGTLPWNGGICCNANPPCFRYPTEGAPGVGNFKSKIVAAFLFSDARKLLLSYOKDT 120
 DB 61 MPSAGTLPWNGGICCNANPPCFRYPTEGAPGVGNFKSKIVAAFLFSDARKLLLSYOKDT 120
 QY 121 SMKDMRKVLRTLQOIKKSSNMLKLDPLVDNTEFSGFLYHNLSPKSTVDKMLRADYILH 180
 DB 121 SMKDMRKVLRTLQOIKKSSNMLKLDPLVDNTEFSGFLYHNLSPKSTVDKMLRADYILH 180
 QY 181 KVFLOGYQLHLTSLCNGSKSEEMIDLGQVSELCGLPKKELAAERVLRSNMIDLPIL 240
 DB 181 KVFLOGYQLHLTSLCNGSKSEEMIDLGQVSELCGLPREKLAERVLRSNMIDLPIL 240
 QY 241 RLTNSTSPPPSKELAEATKTLTSLGTLAQLFMSRMSDKROEVMFLTNVSSSSSTQI 300
 DB 241 RLTNSTSPPPSKELAEATKTLTSLGTLAQLFMSRMSDKROEVMFLTNVSSSSSTQI 300
 QY 301 YQAVSRIVCGHPEGGGLIKISLWYEDNNYKALFEGNGSTEDDAETFYDNSTTPYCNIDLMK 360
 DB 301 YQAVSRIVCGHPEGGGLIKISLWYEDNNYKALFEGNGSTEDDAETFYDNSTTPYCNIDLMK 360
 QY 361 NLESPPLSRITMKALKRPLLVGKILYTPDTPATROYMAVFNKTFQELAFHDLDEGMEELS 420
 DB 361 NLESPPLSRITMKALKRPLLVGKILYTPDTPATROYMAVFNKTFQELAFHDLDEGMEELS 420
 QY 421 PKIWFPMENSQEMDLVRLMLDSRDNDHFEQDLGIDMTADQIVAFIAKHPEDVOSSNGS 480
 DB 421 PKIWFPMENSQEMDLVRLMLDSRDNDHFEQDLGIDMTADQIVAFIAKHPEDVOSSNGS 480
 QY 481 VYTWREAFNETNOAIRTISRMECYNLKLEPIATIEVHLINKSMILLDERKFAAGIVTFG 540
 DB 481 VYTWREAFNETNOAIRTISRMECYNLKLEPIATIEVHLINKSMILLDERKFAAGIVTFG 540
 QY 541 ITPGSTELEPHHYKTRMDIDNVERTNKTKDGYWPGPADPFEDMRKYWGGAVALQDVV 600
 DB 541 ITPGSTELEPHHYKTRMDIDNVERTNKTKDGYWPGPADPFEDMRKYWGGAVALQDVV 600
 QY 601 EDAIIRVLGTGTEKKTGVYQDMPCYVDIDILRWASRMPFLTLAWIYSAVVIKGIY 660
 DB 601 EDAIIRVLGTGTEKKTGVYQDMPCYVDIDILRWASRMPFLTLAWIYSAVVIKGIY 660

```

Db 601 EQAIIIRVLGTEKKTGYVMQMPYVVDITFLRVMSRSMPLFMTLAMIYSVAVILKGIY 660
QY 661 YEKEARKETMRIMGLDINSILWFSEWISLIPLVASALLVILKGLNLPIYSPVAVFV 720
Db 661 YEKEARKETMRIMGLDINSILWFSEWISLIPLVASALLVILKGLNLPIYSPVAVFV 720
QY 721 FLTSVAVAVTIILOCFILISFLFSRANLAACGGIIYFTILYLPVLCVAMODVYGFLLKIFAS 780
Db 721 FLTSVAVAVTIILOCFILISFLFSRANLAACGGIIYFTILYLPVLCVAMODVYGFLLKIFAS 780
QY 781 LLSPAVFGFCYFALFEEOGIVQWMDLFSPEBEGFNITTSISMLFDTPLGYVMTW 840
Db 781 LLSPAVFGFCYFALFEEOGIVQWMDLFSPEBEGFNITTSISMLFDTPLGYVMTW 840
QY 841 YIEAVFPQOYGIIPRWPYEPCTKSYWFGESDEKSHPGSNOKRMEICMEEPHILKIGVS 900
Db 841 YIEAVFPQOYGIIPRWPYEPCTKSYWFGESDEKSHPGSNOKRMEICMEEPHILKIGVS 900
QY 901 IONLVKVRDGMKAVDGLALNFEBOITTSFLGHNGAGKTTMSILNGLFPPTSGTAYTL 960
Db 901 IONLVKVRDGMKAVDGLALNFEBOITTSFLGHNGAGKTTMSILNGLFPPTSGTAYTL 960
QY 961 GKDISEMTIRONLGVCPQHNVLFDMLTYEBHTWYFARLKGISEKIVKAEQMALDVG 1020
Db 961 GKDISEMTIRONLGVCPQHNVLFDMLTYEBHTWYFARLKGISEKIVKAEQMALDVG 1020
QY 1021 LPSSKLKSKTSIQLSGQMOKRLSVALAEVGSKVYILDEPTAGVDPYSRRRIITWELLKRYO 1080
Db 1021 LPSSKLKSKTSIQLSGQMOKRLSVALAEVGSKVYILDEPTAGVDPYSRRRIITWELLKRYO 1080
QY 1081 GRTIITLHHMDEADVIGDRIATISHRKLCQVSSSLFKMQLCTGYTLVLYKKDVSSIS 1140
Db 1081 GRTIITLHHMDEADVIGDRIATISHRKLCQVSSSLFKMQLCTGYTLVLYKKDVSSIS 1140
QY 1141 SCRNSSTSVYSLKEDSVSSSDAGLGSDESDTLTIVSAISNLIRKHVSARLVEI 1200
Db 1141 SCRNSSTSVYSLKEDSVSSSDAGLGSDESDTLTIVSAISNLIRKHVSARLVEI 1200
QY 1201 GHELTLYVLPYPAAKEGAFVELFHEIDRLSDLGISSTGISETTLEIFLKVAEESGVDAE 1260
Db 1201 GHELTLYVLPYPAAKEGAFVELFHEIDRLSDLGISSTGISETTLEIFLKVAEESGVDAE 1260
QY 1261 TSDGTLPARRNRRAFGKQSCLRPFTEDDAADPNDSDIDPESRETDLLSGMGKGSYQVK 1320
Db 1261 TSDGTLPARRNRRAFGKQSCLRPFTEDDAADPNDSDIDPESRETDLLSGMGKGSYQVK 1320
QY 1321 GMRITQOQFVALLMKRLLIARRSRKGFQAQIYLPVAVFCIALVFSILVPEFGKYSLEIO 1380
Db 1321 GMRITQOQFVALLMKRLLIARRSRKGFQAQIYLPVAVFCIALVFSILVPEFGKYSLEIO 1380
QY 1381 PMWYNBOYTFVSNDAPEDTGTLELINALTKDGFGRCHMEGNIPDPTPCQAGEEWTAP 1440
Db 1381 PMWYNBOYTFVSNDAPEDTGTLELINALTKDGFGRCHMEGNIPDPTPCQAGEEWTAP 1440
QY 1441 VPQTIMDLFQNGWMTQNPSPACQSSDKIKKMLPVCPPGAGGLPPPKOKNTADIILODL 1500
Db 1441 VPQTIMDLFQNGWMTQNPSPACQSSDKIKKMLPVCPPGAGGLPPPKOKNTADIILODL 1500
QY 1501 TGRNISGYLVKTYVQIIAASLKNKIVWNEFRYGFSLGVSNTOALPQOEVDATKQKK 1560
Db 1501 TGRNISGYLVKTYVQIIAASLKNKIVWNEFRYGFSLGVSNTOALPQOEVDATKQKK 1560
QY 1561 HLKLAQSSADRFNLNGLRFMTGLDTRNNVKKWFNNKGNHAISSFLANTINNALIRANLOK 1620
Db 1561 HLKLAQSSADRFNLNGLRFMTGLDTRNNVKKWFNNKGNHAISSFLANTINNALIRANLOK 1620
QY 1621 GENPSHYGITAENHPLNLTQOOLSEVALMTTSDVLVSCVIFAFASFPVAFVFLIOER 1680
Db 1621 GENPSHYGITAENHPLNLTQOOLSEVALMTTSDVLVSCVIFAFASFPVAFVFLIOER 1680
QY 1681 VSKAKHLOFISGVKPYIYWLSNFVDMCNVYVPAVLVIIIFICQOKSYSSSTMLPVIAL 1740
Db 1681 VSKAKHLOFISGVKPYIYWLSNFVDMCNVYVPAVLVIIIFICQOKSYSSSTMLPVIAL 1740

```

```

QY 1741 LLLLYGMSITPLMPASFVEKIPSTAVVLTSVNLFTIGNGSVATEVLEETDNKLNNIN 1800
Db 1741 LLLLYGMSITPLMPASFVEKIPSTAVVLTSVNLFTIGNGSVATEVLEETDNKLNNIN 1800
QY 1801 DILKSVFLIPPHFCLGRLLIDMVKNQAMADALEFGENRRVSPLSMDLVGRNLFAMAVBG 1860
Db 1801 DILKSVFLIPPHFCLGRLLIDMVKNQAMADALEFGENRRVSPLSMDLVGRNLFAMAVBG 1860
QY 1861 VVEFLITVLIOYRFFIIPRPYNAKLSPLNDEDEVDREORRIIDGGGONDILEIKELTETI 1920
Db 1861 VVEFLITVLIOYRFFIIPRPYNAKLSPLNDEDEVDREORRIIDGGGONDILEIKELTETI 1920
QY 1921 YRRKRRKPAVDICVIGIPPECFCGLLVGNAGKSSSTEKMLTGDTTVTYRGDAFLNKNSILSN 1980
Db 1921 YRRKRRKPAVDICVIGIPPECFCGLLVGNAGKSSSTEKMLTGDTTVTYRGDAFLNKNSILSN 1980
QY 1981 IHEVHONMGVCPQPDALTTELITGHEHEFFALLRGVPEKEVGVGEMARLKLGLVKGK 2040
Db 1981 IHEVHONMGVCPQPDALTTELITGHEHEFFALLRGVPEKEVGVGEMARLKLGLVKGK 2040
QY 2041 YAGNYSGNKRKRLSTAMALIGCPPVFLDEPTTGMPKARRELMNCALSVKEGSRVLT 2100
Db 2041 YAGNYSGNKRKRLSTAMALIGCPPVFLDEPTTGMPKARRELMNCALSVKEGSRVLT 2100
QY 2101 SHSMECEALCTRAIIVNCRFCIJSVQHLKRRSGDITTVIRIAGSNPDLKPVQDEFG 2160
Db 2101 SHSMECEALCTRAIIVNCRFCIJSVQHLKRRSGDITTVIRIAGSNPDLKPVQDEFG 2160
QY 2161 LAFPGSVLKEKHNMLQYQPLSSLSLARIFSTLSGSKRLHIEBYSQOTLLDQVFNF 2220
Db 2161 LAFPGSVLKEKHNMLQYQPLSSLSLARIFSTLSGSKRLHIEBYSQOTLLDQVFNF 2220
QY 2221 AKQSDDDHLKDLSLHKQTVADVAVLTSFLODEKYESYV 2261
Db 2221 AKQSDDDHLKDLSLHKQTVADVAVLTSFLODEKYESYV 2261

RESULT 8
AAB71749
ID AAB71749 standard; protein; 2261 AA.
XX
AC AAB71749;
XX
DT 17-MAY-2001 (first entry)
XX
DE Human ABC1 protein.
XX
KW High density lipoprotein-cholesterol; HDL-C; cardiovascular; ABC1.
XX
OS Homo sapiens.
XX
PN M0200115676-A2.
XX
PD 08-MAR-2001.
XX
PF 01-SEP-2000; 2000MO-1B01492.
XX
PR 01-SEP-1999; 99US-0151977.
PR 15-MAR-2000; 2000US-0526193.
PR 23-JUN-2000; 2000US-0213958.
XX
PA (UYBR-) UNIV BRITISH COLUMBIA.
PA (XENO-) XENON GENETICS INC.
XX
PI Hayden MR, Brooks-Wilson AR, Pimstone SN, Clee SK;
XX
DR WPL; 2001-244356/25.
XX
PT Treating a lower than normal high density lipoprotein-cholesterol
PT (HDL-C) level, a higher than normal triglyceride level, or a
PT cardiovascular disease, by administering a compound that modulates LXR-
PT or RXR-mediated transcriptional activity -

```

XX Claim 16; Fig 2; 317pp; English.

CC The present invention relates to a method for treating a patient
CC diagnosed as having a lower than normal high density
CC lipoprotein-cholesterol (HDL-C) level, a higher than normal
CC triglyceride level, or a cardiovascular disease, involving
CC administering a compound that modulates LXR- or RXR-mediated
CC transcriptional activity or ABC1 expression or activity.
CC The LXR gene product may be used in an assay to identify
CC compounds useful for the treatment of a disease or condition selected a
CC lower than normal HDL cholesterol level, a higher than normal
CC triglyceride level, and a cardiovascular disease.

XX Sequence 2261 AA;

Query Match 99.9%; Score 11789; DB 22; Length 2261;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 2258; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 MACWPOLRLMKNTFRRCOTCOLLEAVAMPLEFILLISVRLSYRPEYOECHFPNKA 60
DB 1 MACWPOLRLMKNTFRRCOTCOLLEAVAMPLEFILLISVRLSYRPEYOECHFPNKA 60
QY 61 MESAGTLPWVGILICANNPCFRYPPTGEPAGVGNFNKSIARLSPDARRLLYSOKDT 120
DB 61 MESAGTLPWVGILICANNPCFRYPPTGEPAGVGNFNKSIARLSPDARRLLYSOKDT 120
QY 121 SKMDKRVLRITLQOIKKSSSNLKLODFLVNDNETFSGFLYHNLSPKSTYDKMLRADVILH 180
DB 121 SKMDKRVLRITLQOIKKSSSNLKLODFLVNDNETFSGFLYHNLSPKSTYDKMLRADVILH 180
QY 181 KYVLOGYOLHLSLNGSGSSEMIOLGDOEVELGCLPPEKLAARVLRSMNDILKPL 240
DB 181 KYVLOGYOLHLSLNGSGSSEMIOLGDOEVELGCLPPEKLAARVLRSMNDILKPL 240
QY 241 RLINSTSPPEKELATRTLLHSLGTLAOLEFSMRSWSDMRQEVFLTNVSSSSSTOI 300
DB 241 RLINSTSPPEKELATRTLLHSLGTLAOLEFSMRSWSDMRQEVFLTNVSSSSSTOI 300
QY 301 YQAVSTIVCGHPPGGGLIKSLNMYEDNNYKALFGNGTEDEAFETFDNSTPYCNDLAK 360
DB 301 YQAVSTIVCGHPPGGGLIKSLNMYEDNNYKALFGNGTEDEAFETFDNSTPYCNDLAK 360
QY 361 NLESSPLSRITKALPPLVGLTYPDPATROVAENKPTFOELAVHDELGMEELIS 420
DB 361 NLESSPLSRITKALPPLVGLTYPDPATROVAENKPTFOELAVHDELGMEELIS 420
QY 421 PKITFEMENSQENDLVRMLDSRDNDHFWEQOLGDMTADIVAFIAKHPEDVOSSNGS 480
DB 421 PKITFEMENSQENDLVRMLDSRDNDHFWEQOLGDMTADIVAFIAKHPEDVOSSNGS 480
QY 481 VYTWREAFNETNOAIRTISRFMECVNLKLEPLATEVWLINKSMELDERKFPAGIVFTG 540
DB 481 VYTWREAFNETNOAIRTISRFMECVNLKLEPLATEVWLINKSMELDERKFPAGIVFTG 540
QY 541 ITFGSTIELPHVYKRTIMDDNVERTNKIKDGYWDPGRADPEDRMRYWGGATYQDYY 600
DB 541 ITFGSTIELPHVYKRTIMDDNVERTNKIKDGYWDPGRADPEDRMRYWGGATYQDYY 600
QY 601 EOAIIIRVLGTERTKGTGYMOQMPYPCYVDIPLRVMSRMPLEMTLAWISVAIIKGIY 660
DB 601 EOAIIIRVLGTERTKGTGYMOQMPYPCYVDIPLRVMSRMPLEMTLAWISVAIIKGIY 660
QY 661 YEKERARLKEFMRIKGLDNLILMFSSILPILVSGGLVYLKIGNLLPYSDPSVVEY 720
DB 661 YEKERARLKEFMRIKGLDNLILMFSSILPILVSGGLVYLKIGNLLPYSDPSVVEY 720
QY 721 FLISFAVNTLLQOFLISTESSRANLAAACGIIYFTLYPLPYLCVAMODVGTGKTIFAS 780
DB 721 FLISFAVNTLLQOFLISTESSRANLAAACGIIYFTLYPLPYLCVAMODVGTGKTIFAS 780
QY 781 LLSVAFGFCGEYFALFEEOGIGVQMDNLFESPVEDGENTLTSISMMLPDTFLYGVMTW 840

DB 781 LLSVAFGFCGEYFALFEEOGIGVQMDNLFESPVEDGENTLTSISMMLPDTFLYGVMTW 840
QY 841 YIEAFVPGGYGIRPPWYFPCTKSWFCEESDEKSHGSGNOKRMSEICMEEPHKLIGVS 900
DB 841 YIEAFVPGGYGIRPPWYFPCTKSWFCEESDEKSHGSGNOKRMSEICMEEPHKLIGVS 900
QY 901 IONLVKRYRGMKVAVDGLALNFEGOITSFLHNGAGKTTTMSILTGLFPPTSSTAYIL 960
DB 901 IONLVKRYRGMKVAVDGLALNFEGOITSFLHNGAGKTTTMSILTGLFPPTSSTAYIL 960
QY 961 GKDIRSEMTSTIRQNLGVCPQHNVLFDMLTVEEHIFEARLKLSEKHVKAEMQALDVG 1020
DB 961 GKDIRSEMTSTIRQNLGVCPQHNVLFDMLTVEEHIFEARLKLSEKHVKAEMQALDVG 1020
QY 1021 LPSSKLKSKTSQSLSGMORLKVAVLAVGSKVYLDEPTAGVDPSRRGIMELLKRYR 1080
DB 1021 LPSSKLKSKTSQSLSGMORLKVAVLAVGSKVYLDEPTAGVDPSRRGIMELLKRYR 1080
QY 1081 GRTIILSTHMDADVLDRIATISHGKLCVGSLSFLKNOGTGYVLTLYKKDVESLS 1140
DB 1081 GRTIILSTHMDADVLDRIATISHGKLCVGSLSFLKNOGTGYVLTLYKKDVESLS 1140
QY 1141 SCRNSSSTVSYLKKEDSVSSSDAGLGSDESDTLTIDVSAISNLIRKAVSEARLYEDI 1200
DB 1141 SCRNSSSTVSYLKKEDSVSSSDAGLGSDESDTLTIDVSAISNLIRKAVSEARLYEDI 1200
QY 1201 GHETLYVLYPEAAKEGAFVELFHEIDRLSDLTSSYGISSTTLEELFLVYAESGVDAE 1260
DB 1201 GHETLYVLYPEAAKEGAFVELFHEIDRLSDLTSSYGISSTTLEELFLVYAESGVDAE 1260
QY 1261 TSDGTLPARRRNRRAFGDKOSCLRPFTEDDADPNDSDIDEPSERTDLSGMDGKSYQV 1320
DB 1261 TSDGTLPARRRNRRAFGDKOSCLRPFTEDDADPNDSDIDEPSERTDLSGMDGKSYQV 1320
QY 1321 GSKLTQOQFVALLMKRLIARSRKGFPAQVLPVAVYCALAVSLVPPGKYPSEIQ 1380
DB 1321 GSKLTQOQFVALLMKRLIARSRKGFPAQVLPVAVYCALAVSLVPPGKYPSEIQ 1380
QY 1381 PMWVNEQYTFVSNADPEDTGTLELINALTKPDGFCGTCMEGNPFPDTPCOAGEEMTAP 1440
DB 1381 PMWVNEQYTFVSNADPEDTGTLELINALTKPDGFCGTCMEGNPFPDTPCOAGEEMTAP 1440
QY 1441 VPQITMDLFQNGNMTQNPSPACQSSDKIKKMLPVCPPGAGGLPPQPKONTADILQDL 1500
DB 1441 VPQITMDLFQNGNMTQNPSPACQSSDKIKKMLPVCPPGAGGLPPQPKONTADILQDL 1500
QY 1501 TGRNISTDLVKTYYOITAKSLKNKIWNEFRYGGFSLGVSTQALPSPQEVNDAIKQMK 1560
DB 1501 TGRNISTDLVKTYYOITAKSLKNKIWNEFRYGGFSLGVSTQALPSPQEVNDAIKQMK 1560
QY 1561 HUKLAKDSSADREFLNSLGRPMTGLDTRNNVAVYFNNKGMHAISFLVYINNAIRANLQK 1620
DB 1561 HUKLAKDSSADREFLNSLGRPMTGLDTRNNVAVYFNNKGMHAISFLVYINNAIRANLQK 1620
QY 1621 GENPSHYGTTAFNHLNLTKOOLSEVALMTTSVDVLSICVIFAMSEVPASVVFLLQER 1680
DB 1621 GENPSHYGTTAFNHLNLTKOOLSEVALMTTSVDVLSICVIFAMSEVPASVVFLLQER 1680
QY 1681 VSKAKHLOPISGVKRYIYWLNSFVNDKCNVYVPAVLVILITCQOKSYVSTNLPIVAL 1740
DB 1681 VSKAKHLOPISGVKRYIYWLNSFVNDKCNVYVPAVLVILITCQOKSYVSTNLPIVAL 1740
QY 1741 LLLLGMSITPLMPYSPASFVKIPSTAYVVLVSVNLFIGINSVATFVLELTDKLNIN 1800
DB 1741 LLLLGMSITPLMPYSPASFVKIPSTAYVVLVSVNLFIGINSVATFVLELTDKLNIN 1800
QY 1801 DILKSVFLIFPHCLGRLIDNVKQAMADALERFGENRFPVPSLWDLVGRNLEMAVEG 1860
DB 1801 DILKSVFLIFPHCLGRLIDNVKQAMADALERFGENRFPVPSLWDLVGRNLEMAVEG 1860
QY 1861 VVEFLITVLIQRFPIRRPVPNAKLSPLNDEDEVDVRRRQRTLDGSGGNDILEIKELKI 1920
DB 1861 VVEFLITVLIQRFPIRRPVPNAKLSPLNDEDEVDVRRRQRTLDGSGGNDILEIKELKI 1920

Db 1861 VVFELITVLIOYRFFIRPPVNAKLSPINDEDEYRERKORTLDGGGNDILEIKELTKI 1920
 QY 1921 YRRKKRPADVRLICVIGPEGCEGLLVGACAKSSTFKMLTGDPTVTNRGDAFLINKNSILSN 1980
 Db 1921 YRRKKRPADVRLICVIGPEGCEGLLVGACAKSSTFKMLTGDPTVTNRGDAFLINKNSILSN 1980
 QY 1981 IHEVHONMGYCGOPAFITELLTGREHVEFFALLRGVPEKEKVGEMAIRKLGIVKYEK 2040
 Db 1981 IHEVHONMGYCGOPAFITELLTGREHVEFFALLRGVPEKEKVGEMAIRKLGIVKYEK 2040
 QY 2041 YAGNYSGGKRRKLTAMALIGGPVVFDEPTTGMDPKARRPLMNCALSYVEKRSVLT 2100
 Db 2041 YAGNYSGGKRRKLTAMALIGGPVVFDEPTTGMDPKARRPLMNCALSYVEKRSVLT 2100
 QY 2101 SHSMECEALCTRAIMVNGRRCLGSVOHLKNRFGDYTTIVRIAGSNPDLKPVODFFG 2160
 Db 2101 SHSMECEALCTRAIMVNGRRCLGSVOHLKNRFGDYTTIVRIAGSNPDLKPVODFFG 2160
 QY 2161 IAFPGSVLKEKRRMLQYQPLSSLSLARIFSLSQSKKRRIHEDYSVQTTLDQVEVNF 2220
 Db 2161 IAFPGSVLKEKRRMLQYQPLSSLSLARIFSLSQSKKRRIHEDYSVQTTLDQVEVNF 2220
 QY 2221 AKDQSDDDLKDLSLHKNOTVVDVAVLTSFLQDEKVKESYV 2261
 Db 2221 AKDQSDDDLKDLSLHKNOTVVDVAVLTSFLQDEKVKESYV 2261
 RESULT 9
 AAB31363
 ID AAB31363 standard; Protein: 2261 AA.
 XX AAB31363;
 AC
 DT 20-Apr-2001 (first entry)
 XX
 DE Amino acid sequence of ABC1 polypeptide from Tangier disease patient.
 XX
 KW Human: adenosine triphosphate binding cassette protein 1; ABC1;
 KW apolipoprotein-mediated mobilisation; cholesterol; Tangier disease;
 KW chromosome 9q22-9q31; heart disease; hypercholesterolemia;
 KW atherosclerosis; cholesterol transport.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 587
 FT /note= "this is changed from Arg to Trp in Tangier
 disease"
 XX
 PN WO200078972-A2.
 XX
 PD 28-Dec-2000.
 XX
 PF 16-JUN-2000; 2000WO-US16765.
 XX
 PR 18-JUN-1999; 9905-0140264.
 PR 14-SEP-1999; 9905-0153872.
 PR 19-NOV-1999; 9905-016573.
 PA (CVTH-) CV THERAPEUTICS INC.
 PI
 PI Lawn RM, Wade D, Garvin M;
 XX
 DR WPI; 2001-137812/14.
 XX
 PT Adenosine triphosphate (ATP) binding cassette (ABC) polynucleotide,
 PT useful for the development of agents for the treatment of heart disease
 PT and other disorders associated with hypercholesterolemia and
 PT atherosclerosis -
 XX
 PS Disclosure; Page 176-191; 215pp; English.
 XX
 CC The present sequence represents a human adenosine triphosphate (ATP)

CC binding cassette protein (ABC) 1 polypeptide, and is isolated from
 CC a Tangier disease patient. ABC1 resides in cell membranes and utilises
 CC ATP hydrolysis to transport a wide variety of substrates across the
 CC plasma membrane. ABC1 is a pivotal protein in the apolipoprotein-mediated
 CC mobilisation of intracellular cholesterol stores. ABC1 is defective in
 CC Tangier disease, a genetic disorder characterised by abnormal
 CC HDL-cholesterol metabolism. The ABC1 gene is localised to chromosome
 CC 9q22-9q31. The ABC1 genes and proteins are useful for developing
 CC pharmaceutical agents for the treatment of heart disease and other
 CC disorders associated with hypercholesterolemia and atherosclerosis. The
 CC genes are useful for developing screening assays to screen for compounds
 CC that regulate the expression of genes associated with cholesterol
 CC transport. The genes and proteins are also useful for cholesterol
 CC as diagnostic indicators of cardiovascular disease and other disorders
 CC associated with hypercholesterolemia.
 CC
 XX
 SO Sequence 2261 AA:
 Query Match 99.9%; Score 11789; DB 22; Length 2261;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2260; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MACWPQLRLIMKNTLFRKRTQQLLEVAWPLFLLISVRLSPYEQHECHFPNKA 60
 Db 1 MACWPQLRLIMKNTLFRKRTQQLLEVAWPLFLLISVRLSPYEQHECHFPNKA 60
 QY 61 MPAGTLPMVOGIIICANNPCFRYPGEPGVGVGNFNSIARLFSARLLLSOKDT 120
 Db 61 MPAGTLPMVOGIIICANNPCFRYPGEPGVGVGNFNSIARLFSARLLLSOKDT 120
 QY 121 SMKDMKRVLTLOQIKKSSSNLKLQDFLVNDETFSGLVHNLSPKSTYDKMLRADVILH 180
 Db 121 SMKDMKRVLTLOQIKKSSSNLKLQDFLVNDETFSGLVHNLSPKSTYDKMLRADVILH 180
 QY 181 KYFLOGYQLHLTSLCNGSSEEMIQDQEVSELGLPEKELAAERYLRNSMDILKPTL 240
 Db 181 KYFLOGYQLHLTSLCNGSSEEMIQDQEVSELGLPEKELAAERYLRNSMDILKPTL 240
 QY 241 RITNSTSPPEKLEATRTLLHSIGTLAQELFSKRSMDMQEWMPLTNNSSSSSTOI 300
 Db 241 RITNSTSPPEKLEATRTLLHSIGTLAQELFSKRSMDMQEWMPLTNNSSSSSTOI 300
 QY 301 YQAVSRIVCGHPEGGLKIKSLNMYEDNNYKALFEGNGTEDEAETFYDNSTPYCNDLAK 360
 Db 301 YQAVSRIVCGHPEGGLKIKSLNMYEDNNYKALFEGNGTEDEAETFYDNSTPYCNDLAK 360
 QY 361 NLESSLSTIIRKALKPLLVGKILYTPPTPATROYMAEVNKTFOELAVFHDLGEMWELIS 420
 Db 361 NLESSLSTIIRKALKPLLVGKILYTPPTPATROYMAEVNKTFOELAVFHDLGEMWELIS 420
 QY 421 PKIWTFMENSQEMDLVRMLLDSRDNDHFMEQQLDGLMTADIDYAFIAKHPEDVOSSNGS 480
 Db 421 PKIWTFMENSQEMDLVRMLLDSRDNDHFMEQQLDGLMTADIDYAFIAKHPEDVOSSNGS 480
 QY 481 VYTWREAFNEINQALRTISREMECVNLKLEPIATEVWLIKSMELDERKFMAGIYFTG 540
 Db 481 VYTWREAFNEINQALRTISREMECVNLKLEPIATEVWLIKSMELDERKFMAGIYFTG 540
 QY 541 ITPGSIELPHNHYKIKRNDIDNVEFTNKIKQGYWDCGRADPFEDMRYVMGCAFYLDVV 600
 Db 541 ITPGSIELPHNHYKIKRNDIDNVEFTNKIKQGYWDCGRADPFEDMRYVMGCAFYLDVV 600
 QY 601 EQATIRVLGTGFEKKTGYVMQMPYCYVDJLELRWMSRMLPMTLAMIYSVAATIGIV 660
 Db 601 EQATIRVLGTGFEKKTGYVMQMPYCYVDJLELRWMSRMLPMTLAMIYSVAATIGIV 660
 QY 661 YEKEARKLETKRIMGIDNSILFMSWISLILPLVSAGLIVIIKLGNIPLPYSPSVYFV 720
 Db 661 YEKEARKLETKRIMGIDNSILFMSWISLILPLVSAGLIVIIKLGNIPLPYSPSVYFV 720
 QY 721 FLSTAVAVVTIIQCLISTLFSRANLAACGGIIFTYLYLPVLCVAMQDYVGTFLKIFAS 780
 Db 721 FLSTAVAVVTIIQCLISTLFSRANLAACGGIIFTYLYLPVLCVAMQDYVGTFLKIFAS 780

QY 781 LLSPAFGGCEYFALFEEOGIGVOMDLFESPVEDGFLNLTTSISMALFDTLYGVMTW 840
 |||||
 Db 781 LLSPAFGGCEYFALFEEOGIGVOMDLFESPVEDGFLNLTTSISMALFDTLYGVMTW 840
 QY 841 YLEAVFPGQYGTIPRWYFPCTKSYNFGESDEKSHPGSNOKRMSTICMEEPHTLKLGS 900
 |||||
 Db 841 YLEAVFPGQYGTIPRWYFPCTKSYNFGESDEKSHPGSNOKRMSTICMEEPHTLKLGS 900
 QY 901 IONLVKVRDGMKVAVDGALALNFYEGQITSPFHNGACKTTMTSLTGLPPTSGTAIIL 960
 |||||
 Db 901 IONLVKVRDGMKVAVDGALALNFYEGQITSPFHNGACKTTMTSLTGLPPTSGTAIIL 960
 QY 961 GKDISEMSTIRONLVCVPOHNVLFDMLTVEEHIMFYARLKLSEKHAKEMEQMALDVG 1020
 |||||
 Db 961 GKDISEMSTIRONLVCVPOHNVLFDMLTVEEHIMFYARLKLSEKHAKEMEQMALDVG 1020
 QY 1021 LPSSKLKSTSQLSGGMOKLSVALAFVGGSKVVLDEPTAGVDPRSRGIMWELLKRYQ 1080
 |||||
 Db 1021 LPSSKLKSTSQLSGGMOKLSVALAFVGGSKVVLDEPTAGVDPRSRGIMWELLKRYQ 1080
 QY 1081 GRTIILSTHMEADYVLDGRIATISHGLCCVGSFLFKNOLGTGYLTLYKKDVESLS 1140
 |||||
 Db 1081 GRTIILSTHMEADYVLDGRIATISHGLCCVGSFLFKNOLGTGYLTLYKKDVESLS 1140
 QY 1141 SCRNSSSTVYLKEDSVSOSSSDAGLSDHESDPLTIDVSAISMLIRKHVSEARLVEDI 1200
 |||||
 Db 1141 SCRNSSSTVYLKEDSVSOSSSDAGLSDHESDPLTIDVSAISMLIRKHVSEARLVEDI 1200
 QY 1201 GHELYTVLPYEAAGKAGAYELFHEIDRLSDLGSSYGISETTLEEILFKVAESGVDAE 1260
 |||||
 Db 1201 GHELYTVLPYEAAGKAGAYELFHEIDRLSDLGSSYGISETTLEEILFKVAESGVDAE 1260
 QY 1261 TSDGTLPARNRARAGDOSCRLPTEDDAADPNDDIDPESREFDLSGMDGCSYQVK 1320
 |||||
 Db 1261 TSDGTLPARNRARAGDOSCRLPTEDDAADPNDDIDPESREFDLSGMDGCSYQVK 1320
 QY 1321 GSKLTQOQFVALLMKRLLIARSRKGFPAQIYLPVAFVCIALVFSLIYPPGKYSELEQ 1380
 |||||
 Db 1321 GSKLTQOQFVALLMKRLLIARSRKGFPAQIYLPVAFVCIALVFSLIYPPGKYSELEQ 1380
 QY 1381 PMYMEQYTFVSNDAPEDTGTELLNALTKDPFGSTRMEGNIPIPTPCQAGEEWTAP 1440
 |||||
 Db 1381 PMYMEQYTFVSNDAPEDTGTELLNALTKDPFGSTRMEGNIPIPTPCQAGEEWTAP 1440
 QY 1441 VPQTIMDLFONGNMTMOPNSPACOCSSDIKKMLPVCPRGAGLPBPORKONTADIIODL 1500
 |||||
 Db 1441 VPQTIMDLFONGNMTMOPNSPACOCSSDIKKMLPVCPRGAGLPBPORKONTADIIODL 1500
 QY 1501 TGRNIDSLVKTYYOIIAKSLKKNIMVNEFRYGFSLGVSNTQALPSEQEVDALIKQMK 1560
 |||||
 Db 1501 TGRNIDSLVKTYYOIIAKSLKKNIMVNEFRYGFSLGVSNTQALPSEQEVDALIKQMK 1560
 QY 1561 HLKLAKDSSADRFNLISLGRFMTGLDTRNNVKKVFNKKMHAISSTFLVNTINAILRANLQK 1620
 |||||
 Db 1561 HLKLAKDSSADRFNLISLGRFMTGLDTRNNVKKVFNKKMHAISSTFLVNTINAILRANLQK 1620
 QY 1621 GENPSHYGTTAFNHPDLNLTQOOLSEVALMTTSDVLVSICTVFAMSEVPASVFLIOER 1680
 |||||
 Db 1621 GENPSHYGTTAFNHPDLNLTQOOLSEVALMTTSDVLVSICTVFAMSEVPASVFLIOER 1680
 QY 1681 VSKAKHLDFISGVKPVYIWLNSFVWDMCNYYVPATLVIIIFICFOQKSYVSTNPLVAL 1740
 |||||
 Db 1681 VSKAKHLDFISGVKPVYIWLNSFVWDMCNYYVPATLVIIIFICFOQKSYVSTNPLVAL 1740
 QY 1741 LLLLYGKSTPLMTASTVFKIPSTAYVVLTVSVLFIQINSVATFVLELFTDKNLNIN 1800
 |||||
 Db 1741 LLLLYGKSTPLMTASTVFKIPSTAYVVLTVSVLFIQINSVATFVLELFTDKNLNIN 1800
 QY 1801 DLKSVFLIFPFCLGRGLIDVKKQAMADALERGENRFVSPISMDVGNLFLAMAVEG 1860
 |||||
 Db 1801 DLKSVFLIFPFCLGRGLIDVKKQAMADALERGENRFVSPISMDVGNLFLAMAVEG 1860

QY 1861 VFPLLTVLIQYREFIRPRPVNAKLSPANDEDDVRERORILIDGGGONDILEIKELTKI 1920
 |||||
 Db 1861 VFPLLTVLIQYREFIRPRPVNAKLSPANDEDDVRERORILIDGGGONDILEIKELTKI 1920
 QY 1921 YRRKRPAVDRIKICGIPPGECFGLGVNGAKSSTERKMLTGDTTVIRGDAFLKNSTLSN 1980
 |||||
 Db 1921 YRRKRPAVDRIKICGIPPGECFGLGVNGAKSSTERKMLTGDTTVIRGDAFLKNSTLSN 1980
 QY 1981 IHEYHOMMGVCPQPDATTELLTGEHEHFEFALLRGVREKVGVMGMAIRKLGIVYCKE 2040
 |||||
 Db 1981 IHEYHOMMGVCPQPDATTELLTGEHEHFEFALLRGVREKVGVMGMAIRKLGIVYCKE 2040
 QY 2041 YAGNYSGGNKRKLSTAAALIGPPVVFLEDEPTTGMPKARFLMNCALSVKGGRSVLT 2100
 |||||
 Db 2041 YAGNYSGGNKRKLSTAAALIGPPVVFLEDEPTTGMPKARFLMNCALSVKGGRSVLT 2100
 QY 2101 SHSMECEALCTRMALIVNGHFRCLGSVQHLKMPGDSYTIIVRIAGSNPDLKPVODFFG 2160
 |||||
 Db 2101 SHSMECEALCTRMALIVNGHFRCLGSVQHLKMPGDSYTIIVRIAGSNPDLKPVODFFG 2160
 QY 2161 LAFPGSVLKEKRRMLOYQLPSSLSLARLFSTLSOSKRLHIEDVSVQTTLDQVFVNF 2220
 |||||
 Db 2161 LAFPGSVLKEKRRMLOYQLPSSLSLARLFSTLSOSKRLHIEDVSVQTTLDQVFVNF 2220
 QY 2221 AKQSDDDHLKDLSLKRNQTVVDVAVLTSPLODEKVESYV 2261
 |||||
 Db 2221 AKQSDDDHLKDLSLKRNQTVVDVAVLTSPLODEKVESYV 2261
 RESULT 10
 AAB31367
 ID AAB31367 standard; Protein: 2261 AA.
 XX
 AC AAB31367:
 DT 20-APR-2001 (first entry)
 XX
 DE Amino acid sequence of ABC1 polypeptide from Tangier disease patient.
 XX
 KW Human: adenosine triphosphate binding cassette protein 1; ABC1;
 KW apolipoprotein-mediated mobilisation; cholesterol; Tangier disease;
 KW chromosome 9q22-9q31; heart disease; hypercholesterolemia;
 KW atherosclerosis; cholesterol transport.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT MISC-difference 587
 FT PT /note= "this is changed from Arg to Trp in Tangier disease"
 XX
 PN NC0200078971-A2.
 XX
 PD 28-DEC-2000.
 XX
 PF 16-JUN-2000; 2000MO-US16591.
 XX
 PR 18-JUN-1999; 99US-0140264.
 PR 14-SEP-1999; 99US-0153872.
 PR 19-NOV-1999; 99US-0166573.
 XX
 PA (CVTH-) CV THERAPEUTICS INC.
 PA (UNITW) UNIT WASHINGTON.
 PI Lawn RM, Wade D, Oram JF, Garvin M;
 DR WPI; 2001-137811/14.
 DR N-PSDB; AAF24708.
 XX
 PT Adenosine triphosphate (ATP) binding cassette protein (ABC) 1
 PT polynucleotides and polypeptides, useful for treatment of heart disease
 PT and other disorders associated with hypercholesterolemia and
 PT atherosclerosis -

XX PS Claim 28; Page 172-187; 211pp; English.
 CC XX The present sequence represents a human adenosine triphosphate (ATP)
 CC binding cassette protein (ABC) 1 polypeptide, and is isolated from
 CC a Tangier disease patient. ABC1 resides in cell membranes and utilises
 CC ATP hydrolysis to transport a wide variety of substrates across the
 CC plasma membrane. ABC1 is a pivotal protein in the apolipoprotein-mediated
 CC mobilisation of intracellular cholesterol stores. ABC1 is defective in
 CC Tangier disease, a genetic disorder characterised by abnormal
 CC HDL-cholesterol metabolism. The ABC1 gene is localised to chromosome
 CC 9q22-9q31. The ABC1 genes and proteins are useful for developing
 CC pharmaceutical agents for the treatment of heart disease and other
 CC disorders associated with hypercholesterolemia and atherosclerosis. The
 CC genes are useful for developing screening assays to screen for compounds
 CC that regulate the expression of genes associated with cholesterol
 CC transport. The genes and proteins are also useful for are also useful
 CC as diagnostic indicators of cardiovascular disease and other disorders
 CC associated with hypercholesterolemia.
 XX Sequence 2261 AA:
 SQ
 Query Match 99.9%; Score 11789; DB 22; Length 2261;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2260; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 721 FLVFAVAVTIIQCEFLISTLFSRANLAAACGIIYFTLYLPLVLCVAMQDVGFTLKIFAS 780
 DB 721 FLVFAVAVTIIQCEFLISTLFSRANLAAACGIIYFTLYLPLVLCVAMQDVGFTLKIFAS 780
 QY 781 LLSVAGFCGEYALFEEOGIGVQWNLFFESPYEEGFMULTISIMLFDITLYGVMTW 840
 DB 781 LLSVAGFCGEYALFEEOGIGVQWNLFFESPYEEGFMULTISIMLFDITLYGVMTW 840
 QY 841 YIEAVFGQVGIPIPPWVFPCTKSYWFGEEDEKSHPSNOKRMSEICMEEPNHLKIGVS 900
 DB 841 YIEAVFGQVGIPIPPWVFPCTKSYWFGEEDEKSHPSNOKRMSEICMEEPNHLKIGVS 900
 QY 901 IQLNVKVRDGMKAVADGLALNFYEGQITSLFNGAGKTTWISILTGLFPPTSGAVYII 960
 DB 901 IQLNVKVRDGMKAVADGLALNFYEGQITSLFNGAGKTTWISILTGLFPPTSGAVYII 960
 QY 961 GKDIRSEMTIRQLMGVCPQHNVLFQMLTYVEHTWFARLKGLSEKHVKAEMQMLDVC 1020
 DB 961 GKDIRSEMTIRQLMGVCPQHNVLFQMLTYVEHTWFARLKGLSEKHVKAEMQMLDVC 1020
 QY 1021 LPSSKLKSKTSQLSGQMORLSVALAFVGSKVILDEPAGVDYPSRRGIMWELLKTYRQ 1080
 DB 1021 LPSSKLKSKTSQLSGQMORLSVALAFVGSKVILDEPAGVDYPSRRGIMWELLKTYRQ 1080
 QY 1081 GRTIILSTHMDADVIGDRIAIISHGKLCVGSLSPLKQOLCTGYLLVKKDVESSLS 1140
 DB 1081 GRTIILSTHMDADVIGDRIAIISHGKLCVGSLSPLKQOLCTGYLLVKKDVESSLS 1140
 QY 1141 SCRNSSTVSYLKKEDSVSSGSSDAGLGSDESPTLIDVSAISNLIRKVSARLVEDI 1200
 DB 1141 SCRNSSTVSYLKKEDSVSSGSSDAGLGSDESPTLIDVSAISNLIRKVSARLVEDI 1200
 QY 1201 GHELTLYLPEAKEGAFVELFHEIDRLSDLGISYISSETTLEIFLKVAEESGVDAE 1260
 DB 1201 GHELTLYLPEAKEGAFVELFHEIDRLSDLGISYISSETTLEIFLKVAEESGVDAE 1260
 QY 1261 TSDGTLPARRRARAFGKQSLRPFETDDAADPDDSDIDESRETDLSGMDKGSYQAK 1320
 DB 1261 TSDGTLPARRRARAFGKQSLRPFETDDAADPDDSDIDESRETDLSGMDKGSYQAK 1320
 QY 1321 GWKLTQOQFVALLMKRLLIARRSKGFFAOIVLPAVFCIALVFSILVPPFGKYPSLEIQ 1380
 DB 1321 GWKLTQOQFVALLMKRLLIARRSKGFFAOIVLPAVFCIALVFSILVPPFGKYPSLEIQ 1380
 QY 1381 PMAVNEOYTFVSNDAPEDTGTLELNAIATKDPGFGTRCMGNGNIPDTPOAGEEWTTAP 1440
 DB 1381 PMAVNEOYTFVSNDAPEDTGTLELNAIATKDPGFGTRCMGNGNIPDTPOAGEEWTTAP 1440
 QY 1441 VPQITMDLFQNGNWTMGNPSPACCCSSDKIKKMLPVCPPAGGILPPQKQKQNTADILQDI 1500
 DB 1441 VPQITMDLFQNGNWTMGNPSPACCCSSDKIKKMLPVCPPAGGILPPQKQKQNTADILQDI 1500
 QY 1501 TGRNISDYLVKTYVQIIIAKSLKNIWNEFRYGGFSLGVSNTQALPPSEVDNAIKQMK 1560
 DB 1501 TGRNISDYLVKTYVQIIIAKSLKNIWNEFRYGGFSLGVSNTQALPPSEVDNAIKQMK 1560
 QY 1561 HLKIAKSSADRFNLNSLGRNTGIDTFRNNKYVFPNNKGHAISSPLNVLINNALIRANLQK 1620
 DB 1561 HLKIAKSSADRFNLNSLGRNTGIDTFRNNKYVFPNNKGHAISSPLNVLINNALIRANLQK 1620
 QY 1621 GENPSHGITAIFNHPMLTQOQLEVALMTTSVDVLVSIQVIFAMSFVPSFVFLIOER 1680
 DB 1621 GENPSHGITAIFNHPMLTQOQLEVALMTTSVDVLVSIQVIFAMSFVPSFVFLIOER 1680
 QY 1681 VSKAKHLQFISGVKPVLYLWLSNFWDMCNVVPATLVIIIFICFOOKSYVSSTNLPLVAL 1740
 DB 1681 VSKAKHLQFISGVKPVLYLWLSNFWDMCNVVPATLVIIIFICFOOKSYVSSTNLPLVAL 1740
 QY 1741 LLLLYGNSITPIWPAFVFKIPSTAVVLTAVVLTAVVLTAVVLTAVVLTAVVLTAVVLTAVV 1800
 DB 1741 LLLLYGNSITPIWPAFVFKIPSTAVVLTAVVLTAVVLTAVVLTAVVLTAVVLTAVVLTAVV 1800

```

QY 1801 DILKSVFLIFPHFCGLRGLIDMVKNOAMADALERFGENREVSPLSMIDLGNFLPAMAVEG 1860
DB 1801 DILKSVFLIFPHFCGLRGLIDMVKNOAMADALERFGENREVSPLSMIDLGNFLPAMAVEG 1860
QY 1861 VVEFLITVLIQYREFIRPRPVNAKISPLINDEDEDVRRERQRIIDGGGONDILEIKELTKI 1920
DB 1861 VVEFLITVLIQYREFIRPRPVNAKISPLINDEDEDVRRERQRIIDGGGONDILEIKELTKI 1920
QY 1921 YRRKKRPVADRICVGIPIPGECFGLLGVNAGAKSSSTFKMLTSDTTVTTRDARLNNKNSILSN 1980
DB 1921 YRRKKRPVADRICVGIPIPGECFGLLGVNAGAKSSSTFKMLTSDTTVTTRDARLNNKNSILSN 1980
QY 1981 IHEVQNNNGYCPQFPAITELLTGREHVEFFALLRGVPEKEVGKYGEMAIKRLGLVYKEGK 2040
DB 1981 IHEVQNNNGYCPQFPAITELLTGREHVEFFALLRGVPEKEVGKYGEMAIKRLGLVYKEGK 2040
QY 2041 YAGNYSNGKRLSTAMALLGGPPVVELDEPTGTMDPKARRFLMNCALSVYKEGSRVLT 2100
DB 2041 YAGNYSNGKRLSTAMALLGGPPVVELDEPTGTMDPKARRFLMNCALSVYKEGSRVLT 2100
QY 2101 SHSMECEALCTRMAIMVNGRFRCLGSVOHLKNRFGDGYTTIVRAGSNPDLKPYODEFG 2160
DB 2101 SHSMECEALCTRMAIMVNGRFRCLGSVOHLKNRFGDGYTTIVRAGSNPDLKPYODEFG 2160
QY 2161 IAFPGSVLKEKRRNMLQYQLPSSLSLARIFSLSQSKRLHIEDYSVSQTTLDQVFVNF 2220
DB 2161 IAFPGSVLKEKRRNMLQYQLPSSLSLARIFSLSQSKRLHIEDYSVSQTTLDQVFVNF 2220
QY 2221 AKDQSDDDLKDLSTLKNQTVVDVAVLTSFLODEKVKESYV 2261
DB 2221 AKDQSDDDLKDLSTLKNQTVVDVAVLTSFLODEKVKESYV 2261

RESULT 11
ID AAB38111 standard; Protein: 2261 AA.
XX
AC AAB38111;
XX
DT 29-JAN-2001 (first entry)
XX
DE Human ABC1 cholesterol transporter mutant, V771M.
XX
KW Human ABC1 cholesterol transporter; chromosome 9q31;
KW ATP-binding cassette; HDL deficiency disorder; high density lipoprotein;
KW Tangier disease; TD; familial HDL deficiency; FHA; polymorphism;
KW cardiovascular disease; coronary artery disease; coronary restenosis;
KW cerebrovascular disease; peripheral vascular disease;
KW Alzheimer's disease; Niemann-Pick disease; Huntington's disease;
KW X-linked adrenoleukodystrophy; cancer; gene therapy; genetic diagnosis;
KW prognosis; prophylaxis; drug screening; transgenic animal; mutant;
XX
XX Homo sapiens.
XX
XX WO200055318-A2.
XX
XX PD 21-SEP-2000.
XX
XX PF 15-MAR-2000; 2000WO-1B00532.
XX
XX PR 15-MAR-1999; 9905-0124702.
XX PR 08-JUN-1999; 9905-0138048.
XX PR 17-JUN-1999; 9905-0139600.
XX PR 01-SEP-1999; 9905-0151977.
XX
XX PA (UYBR-) UNIT BRITISH COLUMBIA.
XX PA (XENO-) XENON BIORESEARCH INC.
XX PI Hayden MR, Wilson AR, Pimstone SN;
XX
XX WPI; 2000-587528/55.

```

```

PT New ABC1 polypeptide is useful for treating diseases associated with
PT ABC1 biological activity, e.g. Alzheimer's disease, Huntington's
PT disease and cancer.
XX
PS Examples: Page -: 229pp; English.
XX
XX The invention relates to the human ABC1 cholesterol transporter protein
XX (B38082) and to nucleic acid sequences (c69120) which encode it. ABC1 is
XX a member of the ATP-binding cassette (ABC transporter) superfamily of
XX proteins, and plays a crucial role in cholesterol transport, particularly
XX intracellular cholesterol trafficking in monocytes and fibroblasts, being
XX involved in cholesterol efflux from the cell. The gene encoding ABC1 is
XX located on chromosome 9q31, and mutations in this gene are associated
XX with two genetic HDL (high density lipoprotein) deficiency disorders,
XX Tangier disease (TD) and familial HDL deficiency (FHA). These diseases
XX are distinguishable in that TD is an autosomal recessive disorder, while
XX FHA is inherited as an autosomal dominant trait. Low levels of HDL ("good
XX cholesterol") in the blood correlate with a high risk of cardiovascular
XX disease, particularly coronary artery disease, but also cerebrovascular
XX disease, coronary restenosis, and peripheral vascular disease.
XX Conversely, a high level of HDL has protective effects against
XX cardiovascular disease. The invention provides genetic constructs and
XX transgenic cells and non-human animals comprising human ABC1 nucleic
XX acids, and methods of gene therapy for the treatment or prevention of
XX cardiovascular disease comprising the administration of an expression
XX vector encoding ABC1 or an active fragment thereof. The invention also
XX encompasses compounds which mimic ABC1 activity, compounds which
XX stimulate ABC1 expression and methods of screening for such compounds.
XX It further relates to methods for determining whether for a patient has an
XX increased risk for cardiovascular disease due to polymorphisms in the
XX ABC1 gene. Human ABC1 proteins and nucleotides can be used to treat
XX or prevent cardiovascular disease, especially coronary artery disease,
XX cerebrovascular disease, coronary restenosis or peripheral vascular
XX disease. They may also be used in the treatment of diseases associated
XX with ABC1 biological activity, such as Alzheimer's disease, Niemann-Pick
XX disease, Huntington's disease, X-linked adrenoleukodystrophy and cancer.
XX The invention specifically excludes proteins with the exact amino acid
XX sequences of GenBank Accession No: CAA10005.1 and X75926, and the nucleic
XX acid with the exact sequence as GenBank Accession No: AJ012376.1. The
XX present sequence represents a mutant human ABC1 cholesterol transporter
XX associated with an altered cholesterol level and therefore an altered
XX risk of cardiovascular disease.
XX Note: The present sequence is not shown in the specification, but is
XX derived from the native human ABC1 shown on pages 152-157.
XX
XX Sequence 2261 AA:
XX
XX Query Match 99.9%; Score 11786; DB 21; Length 2261;
XX Best Local Similarity 99.8%; Pred. No. 0;
XX Matches 2257; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
QY 1 MACWPQLRLLLMKNLTFERRQTCOLLLEVAMPFLIFILISVRLSYPPYDQHECHFPNKA 60
DB 1 MACWPQLRLLLMKNLTFERRQTCOLLLEVAMPFLIFILISVRLSYPPYDQHECHFPNKA 60
QY 61 MPSAGTLPWVGIIICNANNPCFRRYPTEGAPGVGNFNKSIVARLFSFDRRLLYLSQKDT 120
DB 61 MPSAGTLPWVGIIICNANNPCFRRYPTEGAPGVGNFNKSIVARLFSFDRRLLYLSQKDT 120
QY 121 SMKDMRKVLRTLQOITKSSSNLKLQDFLVNDEPFGFYHNLSLPKSTGVKMLRADYILH 180
DB 121 SMKDMRKVLRTLQOITKSSSNLKLQDFLVNDEPFGFYHNLSLPKSTGVKMLRADYILH 180
QY 181 KVFLOGYOLHLTSLCNGSKSEMIQOLGQVESELGCPKREKLAABEVLRSNMDILKPIIL 240
DB 181 KVFLOGYOLHLTSLCNGSKSEMIQOLGQVESELGCPKREKLAABEVLRSNMDILKPIIL 240
QY 241 RTLNSTSPFSKELAEATKTLHLISGLTLAGLFSMRKSDMRQVYMLTLVNNSSSSSTQI 300
DB 241 RTLNSTSPFSKELAEATKTLHLISGLTLAGLFSMRKSDMRQVYMLTLVNNSSSSSTQI 300
QY 301 YQAVSRIVCGHPGGGLTKIKSLNMYEDNNYKALFGGNTGEDAEFFYDNGSTTPPCNDLMMK 360

```

Db 301 YQAVSRIVCGHEGGGKIKSLNWEEDNNKALFEAGNGTEDEAETFDNSTTPYCNDIMK 360
 QY 361 NLESSPESRIWKALKPLLVGRILVTPDT PATROYMAEVNKTFOELAVFHDLEGMWELS 420
 Db 361 NLESSPESRIWKALKPLLVGRILVTPDT PATROYMAEVNKTFOELAVFHDLEGMWELS 420
 QY 421 PKIWTFMENSOEMDLVRLNLDSDRDHFWEOQDLDMDTADDIYAFLAKHEDVOSSNGS 480
 Db 421 PKIWTFMENSOEMDLVRLNLDSDRDHFWEOQDLDMDTADDIYAFLAKHEDVOSSNGS 480
 QY 481 VYTWBAEFNETNOATRTISREMECVNLKLEPIATEVWLINXSMELDEREFMAGIYFTG 540
 Db 481 VYTWBAEFNETNOATRTISREMECVNLKLEPIATEVWLINXSMELDEREFMAGIYFTG 540
 QY 541 ITPGSELPHVHKYKTRMDIDNVEKTNKIKDGYWDPGRADPEEDMYVWGFAVLODYY 600
 Db 541 ITPGSELPHVHKYKTRMDIDNVEKTNKIKDGYWDPGRADPEEDMYVWGFAVLODYY 600
 QY 601 FOAIIIRVLGTGTEKTVGMQMPYPCYVDJIEFRVMSRSMLEMTLAMITYSVANTIKGIY 660
 Db 601 FOAIIIRVLGTGTEKTVGMQMPYPCYVDJIEFRVMSRSMLEMTLAMITYSVANTIKGIY 660
 QY 661 YEKEARLKETRMIMJIDNSILMFSEJSSILPLVSAGLVILKLGNLIPYSDPSVFEV 720
 Db 661 YEKEARLKETRMIMJIDNSILMFSEJSSILPLVSAGLVILKLGNLIPYSDPSVFEV 720
 QY 721 FLVSAVVTIIIOCLISTLSESRANLAAACGIIYFTLYLPYVCVAMQDYGFTLKIFAS 780
 Db 721 FLVSAVVTIIIOCLISTLSESRANLAAACGIIYFTLYLPYVCVAMQDYGFTLKIFAS 780
 QY 781 LLSPAFGGCEYFALFEQSIGVQWDMNFESPVEEDGFNTTISIMLFPOTILYGMW 840
 Db 781 LLSPAFGGCEYFALFEQSIGVQWDMNFESPVEEDGFNTTISIMLFPOTILYGMW 840
 QY 841 YIEAVFPGQYCIIPRPWYFPCYKSYWGESEDEKSHPSNOKRMEICMEEPHTLKIGVS 900
 Db 841 YIEAVFPGQYCIIPRPWYFPCYKSYWGESEDEKSHPSNOKRMEICMEEPHTLKIGVS 900
 QY 901 IONLVKVRDGMKVAVDGLANFEGQITSPFGHNGAKTTTMSILTLGLFPPISGTAYIL 960
 Db 901 IONLVKVRDGMKVAVDGLANFEGQITSPFGHNGAKTTTMSILTLGLFPPISGTAYIL 960
 QY 961 GKDIRSEMSTIRONLGYCPOHNVLFEDMLTVEEHIMFYARLKGISEKHVAKEMEMALDVG 1020
 Db 961 GKDIRSEMSTIRONLGYCPOHNVLFEDMLTVEEHIMFYARLKGISEKHVAKEMEMALDVG 1020
 QY 1021 LPSKLSKTSQLSGGMQRKLSVALAVGSGKVVITDEPTAGVDPYSRSGIWEILLKYRO 1080
 Db 1021 LPSKLSKTSQLSGGMQRKLSVALAVGSGKVVITDEPTAGVDPYSRSGIWEILLKYRO 1080
 QY 1081 GRITITLTHHMDADVLCGRITAIISHGKLCVGSSTPLKNOLOGVYLTLYKKDVESSLS 1140
 Db 1081 GRITITLTHHMDADVLCGRITAIISHGKLCVGSSTPLKNOLOGVYLTLYKKDVESSLS 1140
 QY 1141 SCRNSSTVSYLKREDSVSSQSDAGLSDHESDTEITDIVSAINLIRKHVSARLVEDI 1200
 Db 1141 SCRNSSTVSYLKREDSVSSQSDAGLSDHESDTEITDIVSAINLIRKHVSARLVEDI 1200
 QY 1201 GHELTJYVLPYEAKEGAFVFLFHEIDRLSDIGISSGISETLLEIFLKVAEESGVDAE 1260
 Db 1201 GHELTJYVLPYEAKEGAFVFLFHEIDRLSDIGISSGISETLLEIFLKVAEESGVDAE 1260
 QY 1261 TSDGTLPARNRRAFGKOSGLRPFTEDDAADPNDSDIDPESRETDILSGMDKGSYOAK 1320
 Db 1261 TSDGTLPARNRRAFGKOSGLRPFTEDDAADPNDSDIDPESRETDILSGMDKGSYOAK 1320
 QY 1321 GWKLTQOQFVALLMKRLLIARSRKGFPAQIVLPAVFCIALVFSILVPECKYPSLEIQ 1380
 Db 1321 GWKLTQOQFVALLMKRLLIARSRKGFPAQIVLPAVFCIALVFSILVPECKYPSLEIQ 1380
 QY 1381 PMWYNEGYTVSNDAPDPTSTELNALTAKDPEGRCHMGKNIPPTPOAGCEEWTTAP 1440
 Db 1381 PMWYNEGYTVSNDAPDPTSTELNALTAKDPEGRCHMGKNIPPTPOAGCEEWTTAP 1440

QY 1441 VPOTIMDLFONGMWTQNPSPACQSSDKIKKMLPVCPPGAGGLPPPOKONTADTLODL 1500
 Db 1441 VPOTIMDLFONGMWTQNPSPACQSSDKIKKMLPVCPPGAGGLPPPOKONTADTLODL 1500
 QY 1501 TGRNISDYLVKTYVOJIIANSKANKIVNEFRYGFSLGVSNTOALPPSOEVNDAIKQMK 1560
 Db 1501 TGRNISDYLVKTYVOJIIANSKANKIVNEFRYGFSLGVSNTOALPPSOEVNDAIKQMK 1560
 QY 1561 HLKLANDSSADRFNLISGRMTGLDTRNNVKNVFNKKGHAISSPLNTVNNAILRANIOK 1620
 Db 1561 HLKLANDSSADRFNLISGRMTGLDTRNNVKNVFNKKGHAISSPLNTVNNAILRANIOK 1620
 QY 1621 GENPSHYGITAENHPLNLTIKOOLSEVALMTSYDVLSICVIFAMSFPVASFVEFLIOER 1680
 Db 1621 GENPSHYGITAENHPLNLTIKOOLSEVALMTSYDVLSICVIFAMSFPVASFVEFLIOER 1680
 QY 1681 VSKAKHLOEISGVKPVYIYWLNSFWDMCNVVPATLVIIIFICFOOKSVYSTNLPVAL 1740
 Db 1681 VSKAKHLOEISGVKPVYIYWLNSFWDMCNVVPATLVIIIFICFOOKSVYSTNLPVAL 1740
 QY 1741 LLLYGMSTITPLMKPASFEFKIPSTAYVYLVTSVNLFTGINSVATFVLELFTDNKLNIN 1800
 Db 1741 LLLYGMSTITPLMKPASFEFKIPSTAYVYLVTSVNLFTGINSVATFVLELFTDNKLNIN 1800
 QY 1801 DILKSVFLIFPHECLGRGLDMVKNQAMADALERGENRFVPLSMDLVGRMLFMAVEG 1860
 Db 1801 DILKSVFLIFPHECLGRGLDMVKNQAMADALERGENRFVPLSMDLVGRMLFMAVEG 1860
 QY 1861 VYFELLTYVLIQYRFIRRPVNAKLSPLNDEDEVDVRRERQRLIDGGGONDILEIKELTKI 1920
 Db 1861 VYFELLTYVLIQYRFIRRPVNAKLSPLNDEDEVDVRRERQRLIDGGGONDILEIKELTKI 1920
 QY 1921 YRRKRKPAVDRLICVGIIPPECFGLGVNAGKSTFEKMLTGTPTTTRGDAPLNKSTILSN 1980
 Db 1921 YRRKRKPAVDRLICVGIIPPECFGLGVNAGKSTFEKMLTGTPTTTRGDAPLNKSTILSN 1980
 QY 1981 IHEVHONMGYCOFPAITELLTGREHVEFFALLRGVPEKVGKVGEMAIRKLGIVKGEK 2040
 Db 1981 IHEVHONMGYCOFPAITELLTGREHVEFFALLRGVPEKVGKVGEMAIRKLGIVKGEK 2040
 QY 2041 YAGNSGGRKRLSTAMALIGRPVVFIDEPTTGMDPKARRIIMCALSVYKEGSSVLT 2100
 Db 2041 YAGNSGGRKRLSTAMALIGRPVVFIDEPTTGMDPKARRIIMCALSVYKEGSSVLT 2100
 QY 2101 SHSMECEALCTRMAIMVNGFRCLGSVOHLKNRFGDGYTIVIRIAGSNPDLKPVQDFPG 2160
 Db 2101 SHSMECEALCTRMAIMVNGFRCLGSVOHLKNRFGDGYTIVIRIAGSNPDLKPVQDFPG 2160
 QY 2161 LAFPGSVYLKEKIRNMLOVLPSSLSLARIFSLISQSKRRLHIEDYSVSQTTLDQVEVNF 2220
 Db 2161 LAFPGSVYLKEKIRNMLOVLPSSLSLARIFSLISQSKRRLHIEDYSVSQTTLDQVEVNF 2220
 QY 2221 AKDQSDDDHLKDLSHKNOYVDAVLTSLFLODEKVKESYV 2261
 Db 2221 AKDQSDDDHLKDLSHKNOYVDAVLTSLFLODEKVKESYV 2261
 RESULT 12
 AAB38114
 ID AAB38114 standard; Protein: 2261 AA.
 XX
 AC AAB38114:
 XX
 DT 29-JAN-2001 (first entry)
 XX
 DE Human ABC1 cholesterol transporter mutant, E1172D.
 XX
 KW Human ABC1 cholesterol transporter; chromosome 9q31;
 KW ATP-binding cassette; HDL deficiency disorder; high density lipoprotein;
 KW Tangier disease; Td; familial HDL deficiency; FNA; polymorphism;
 KW cardiovascular disease; coronary artery disease; coronary restenosis;
 KW cerebrovascular disease; peripheral vascular disease;

KW Alzheimer's disease; Niemann-Pick disease; Huntington's disease;
 KW X-linked adrenoleukodystrophy; cancer; gene therapy; genetic diagnosis;
 KW proinosis; prophylaxis; drug screening; transgenic animal; mutant;
 KW mutant.
 OS Homo sapiens.
 PN W020005318-A2.
 XX 21-SEP-2000.
 PD 15-MAR-2000; 2000WO-1B00532.
 PF 15-MAR-1999; 99US-0124702.
 PR 08-JUN-1999; 99US-0138048.
 PR 17-JUN-1999; 99US-0139600.
 PR 01-SEP-1999; 99US-0151977.
 XX (UYBR-) UNITV BRITISH COLUMBIA.
 PA (XENO-) XENON BIORESEARCH INC.
 XX Hayden MR, Wilson AR, Plimstone SN;
 PI WPI; 2000-587528/55.
 DR New ABC1 polypeptide is useful for treating diseases associated with
 PT ABC1 biological activity, e.g. Alzheimer's disease, Huntington's
 PT disease and cancer.
 PS Examples; Page -: 229pp; English.
 XX The invention relates to the human ABC1 cholesterol transporter protein
 CC (B38082) and to nucleic acid sequences (C69120) which encode it. ABC1 is
 CC a member of the ATP-binding cassette (ABC transporter) superfamily of
 CC proteins, and plays a crucial role in cholesterol transport, particularly
 CC intracellular cholesterol trafficking in monocytes and fibroblasts, being
 CC involved in cholesterol efflux from the cell. The gene encoding ABC1 is
 CC located on chromosome 9q31, and mutations in this gene are associated
 CC with two genetic HDL (high density lipoprotein) deficiency disorders,
 CC Tangier disease (TD) and familial HDL deficiency (FHD). These diseases
 CC are distinguishable in that TD is an autosomal recessive disorder, while
 CC FHD is inherited as an autosomal dominant trait. Low levels of HDL ("good
 CC cholesterol") in the blood correlate with a high risk of cardiovascular
 CC disease, particularly coronary artery disease, but also cerebrovascular
 CC disease, coronary restenosis, and peripheral vascular disease.
 CC Conversely, a high level of HDL has protective effects against
 CC cardiovascular disease. The invention provides genetic constructs and
 CC transgenic cells and non-human animals comprising human ABC1 nucleic
 CC acids, and methods of gene therapy for the treatment or prevention of
 CC cardiovascular disease comprising the administration of an expression
 CC vector encoding ABC1 or an active fragment thereof. The invention also
 CC encompasses compounds which mimic ABC1 activity, compounds which
 CC stimulate ABC1 expression and methods of screening for such compounds.
 CC It further relates to methods for determining whether a patient has an
 CC increased risk for cardiovascular disease due to polymorphisms in the
 CC ABC1 gene. Human ABC1 proteins and nucleotides can be used to treat
 CC or prevent cardiovascular disease, especially coronary artery disease,
 CC cerebrovascular disease, coronary restenosis or peripheral vascular
 CC disease. They may also be used in the treatment of diseases associated
 CC with ABC1 biological activity, such as Alzheimer's disease, Niemann-Pick
 CC disease, Huntington's disease, X-linked adrenoleukodystrophy and cancer.
 CC The invention specifically excludes proteins with the exact amino acid
 CC sequences of Genbank Accession No: CAA10005.1 and X75926, and the nucleic
 CC acid with the exact sequence as Genbank Accession No: AJ012376.1. The
 CC present sequence represents a mutant human ABC1 cholesterol transporter
 CC associated with an altered cholesterol level and therefore an altered
 CC risk of cardiovascular disease.
 CC Note: The present sequence is not shown in the specification, but is
 CC derived from the native human ABC1 shown on pages 152-157.
 XX
 XX
 SQ Sequence 2261 AA;
 Query Match 99.9%; Score 11786; DB 21; Length 2261;

Best Local Similarity 99.8%; Pred. No. 0;
 Matches 2257; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MACHPOLRLMKLNLFRRQOTQQLLEVAMPFLIFILISVLSLSPPEQHECHPNKA 60
 DB 1 MACHPOLRLMKLNLFRRQOTQQLLEVAMPFLIFILISVLSLSPPEQHECHPNKA 60
 QY 61 MPSAGTLPWVGIIICNANNPCFRPPPGAPGVGVGNFNKSTVARLSDSARLLYSQKPT 120
 DB 61 MPSAGTLPWVGIIICNANNPCFRPPPGAPGVGVGNFNKSTVARLSDSARLLYSQKPT 120
 QY 121 SMKDMRKVLRTLOQIKKSSNKLQDFLVNDEFSGFLYHNLSLPRSTYDKMLRADVILA 180
 DB 121 SMKDMRKVLRTLOQIKKSSNKLQDFLVNDEFSGFLYHNLSLPRSTYDKMLRADVILA 180
 QY 181 KVFLOGYQLHITSLSCKNSKSEEMIQLDQDVSELSGIPKPKKLAARVLRSMNDILKPLT 240
 DB 181 KVFLOGYQLHITSLSCKNSKSEEMIQLDQDVSELSGIPKPKKLAARVLRSMNDILKPLT 240
 QY 241 RTLNSTSPPSKELAEATKTLHSLGTLAQLFMSMSMSDMROEVLFTVNSSSSSTQI 300
 DB 241 RTLNSTSPPSKELAEATKTLHSLGTLAQLFMSMSMSDMROEVLFTVNSSSSSTQI 300
 QY 301 YQAVSRIVCGHPEGGLKTKSLNNYEDNNYKALFGNGTEDEAETFDNSTTPYCNDLAK 360
 DB 301 YQAVSRIVCGHPEGGLKTKSLNNYEDNNYKALFGNGTEDEAETFDNSTTPYCNDLAK 360
 QY 361 NLESSPLSRILWKALKPLVGLKILYTPDTPATROVAAEVKTKTQELAVFHDLEGMEELS 420
 DB 361 NLESSPLSRILWKALKPLVGLKILYTPDTPATROVAAEVKTKTQELAVFHDLEGMEELS 420
 QY 421 PKITWPMENSOEMDLVAMLDSRDHFMEOQLDGLMWTQODIVAFIAKHPEDVOSSNGS 480
 DB 421 PKITWPMENSOEMDLVAMLDSRDHFMEOQLDGLMWTQODIVAFIAKHPEDVOSSNGS 480
 QY 481 VYTRAEAFNEINQAIRTISRMECVNLKLEPLATEVWMLINKSMELDERKFMAGIVFTG 540
 DB 481 VYTRAEAFNEINQAIRTISRMECVNLKLEPLATEVWMLINKSMELDERKFMAGIVFTG 540
 QY 541 ITPPSIELPHHVKKIRINDINVERTKIKDGYNDPPRADPPEDMKYVWGGAIVQDYY 600
 DB 541 ITPPSIELPHHVKKIRINDINVERTKIKDGYNDPPRADPPEDMKYVWGGAIVQDYY 600
 QY 601 EQAIRVLTGTEKTKGYVMQMPYCYVDIIFLRVMSRMPLENTLAMIYSVAVIIKGIY 660
 DB 601 EQAIRVLTGTEKTKGYVMQMPYCYVDIIFLRVMSRMPLENTLAMIYSVAVIIKGIY 660
 QY 661 YEKARLKEINRIMGDINSILMFSMTSSILPLLSAGLLVILIKGNLLPYSDPSVVFV 720
 DB 661 YEKARLKEINRIMGDINSILMFSMTSSILPLLSAGLLVILIKGNLLPYSDPSVVFV 720
 QY 721 FLSFAVVTIQCFLISTFSTRANLAAAGGIIYFTLYLPVLCVADQVVGFTLKIFAS 780
 DB 721 FLSFAVVTIQCFLISTFSTRANLAAAGGIIYFTLYLPVLCVADQVVGFTLKIFAS 780
 QY 781 LLSPAFGFGEYFALFEEOGIGVOMDLNLFESPYEEDGFNLTTISMMLDFTFLYGVMTV 840
 DB 781 LLSPAFGFGEYFALFEEOGIGVOMDLNLFESPYEEDGFNLTTISMMLDFTFLYGVMTV 840
 QY 841 YIEAVFPGQYGIIPRWYFPCTKSYWFGESDEKSHPSGNKRMSIEICMEEPHTLKIGVS 900
 DB 841 YIEAVFPGQYGIIPRWYFPCTKSYWFGESDEKSHPSGNKRMSIEICMEEPHTLKIGVS 900
 QY 901 IONLVKVRDGMKAAVVDLALNFEQGTTSFLGHNGAGKTTTMSILGLPPPSGTVAIL 960
 DB 901 IONLVKVRDGMKAAVVDLALNFEQGTTSFLGHNGAGKTTTMSILGLPPPSGTVAIL 960
 QY 961 GKDIRSEMTIRONLGVCPQNNVLFDMLTVEHTMFAARLKGISEKKHVKAEMEMALDVG 1020
 DB 961 GKDIRSEMTIRONLGVCPQNNVLFDMLTVEHTMFAARLKGISEKKHVKAEMEMALDVG 1020
 QY 1021 LPSSKLKSTQSOLSGMOKLSVALAFVGSKVILDEPTAGVDPYSRRIWELLKRYK 1080
 DB 1021 LPSSKLKSTQSOLSGMOKLSVALAFVGSKVILDEPTAGVDPYSRRIWELLKRYK 1080

Dp	1021	LPSSMLKSKTSQOLSGGMQRKLSVALAPVAGSSKYVILDEPAGVADPYSRKGIVELLKTRQ	1080
Qy	1081	GRTTILSTHMHDEADVGDRIAIIISHGKLCOCVSSSEFLKKNQLOCTGYTLTVKRDVSSLS	1140
Dp	1081	GRTTILSTHMHDEADVGDRIAIIISHGKLCOCVSSSEFLKKNQLOCTGYTLTVKRDVSSLS	1140
Qy	1141	SCRNSSSVTVLKKEDSVSSOSSAGISGDSHEDTLLTVSAISNLRKHVSEALVEDI	1200
Dp	1141	SCRNSSSVTVLKKEDSVSSOSSAGISGDSHEDTLLTVSAISNLRKHVSEALVEDI	1200
Qy	1201	GHELTYYLPEAAKEGAFVELFHEIDRLSDLCISSYGISSETTLEETFLKVAEEGVAE	1260
Dp	1201	GHELTYYLPEAAKEGAFVELFHEIDRLSDLCISSYGISSETTLEETFLKVAEEGVAE	1260
Qy	1261	TSDGTLPRARRRRAFGCKOSLRPPTTDDAADPNDSIDIPESREFDLSGMDGKSVYK	1320
Dp	1261	TSDGTLPRARRRRAFGCKOSLRPPTTDDAADPNDSIDIPESREFDLSGMDGKSVYK	1320
Qy	1321	GWKLTOQOQFVALLMKRLIILARRSKGFFAOTVPAFVFCALVFEISLTPPEKYSLELO	1380
Dp	1321	GWKLTOQOQFVALLMKRLIILARRSKGFFAOTVPAFVFCALVFEISLTPPEKYSLELO	1380
Qy	1381	PMWNTNEOYTFVNSNDAPEDTGTLELNLATKDPGEFGRCKEKNP1DUTPCQAGEEWTAP	1440
Dp	1381	PMWNTNEOYTFVNSNDAPEDTGTLELNLATKDPGEFGRCKEKNP1DUTPCQAGEEWTAP	1440
Qy	1441	VPOITMDLFQWGNMTMONPSPACOSSDKIKKILPCPCPAGAGLPPPOKQOMTAD1LODL	1500
Dp	1441	VPOITMDLFQWGNMTMONPSPACOSSDKIKKILPCPCPAGAGLPPPOKQOMTAD1LODL	1500
Qy	1501	TGRNISOVLVYTVQOITAKSLKKNIVWNEFRGGSLSVSNQALPPSOEVNDATIKOMK	1560
Dp	1501	TGRNISOVLVYTVQOITAKSLKKNIVWNEFRGGSLSVSNQALPPSOEVNDATIKOMK	1560
Qy	1561	HLKLAKDOSSADRFANSJGRMTGJDFERNVKNVFNKKGHAISSEFLYNNALIBRANLOK	1620
Dp	1561	HLKLAKDOSSADRFANSJGRMTGJDFERNVKNVFNKKGHAISSEFLYNNALIBRANLOK	1620
Qy	1621	GENPSHGITAFNHPMLTKOOLSEVALMTTSVDVLSICVIFPAMSFPVASHVVELIOER	1680
Dp	1621	GENPSHGITAFNHPMLTKOOLSEVALMTTSVDVLSICVIFPAMSFPVASHVVELIOER	1680
Qy	1681	VSKAKHLQFISGKVPVYVWLSNFWVMQCNVVPATVLIITLFCFOOKSVYSTNLPVAL	1740
Dp	1681	VSKAKHLQFISGKVPVYVWLSNFWVMQCNVVPATVLIITLFCFOOKSVYSTNLPVAL	1740
Qy	1741	LLLLGWSITPLMYPASVEFKIPSTAYVVLTSVNLFIGINGSVATFVELETFDNKLNNIN	1800
Dp	1741	LLLLGWSITPLMYPASVEFKIPSTAYVVLTSVNLFIGINGSVATFVELETFDNKLNNIN	1800
Qy	1801	DLKSVFLFPHPCILGGLIDWKKNQAMADALBERGENRFVSPLSMDLYGRNLFMAVEG	1860
Dp	1801	DLKSVFLFPHPCILGGLIDWKKNQAMADALBERGENRFVSPLSMDLYGRNLFMAVEG	1860
Qy	1861	VVEFLITVLLOYREFITRPREVNAKSLPLNDEDEEDVAREKORILIDGGQNDLIELKELKI	1920
Dp	1861	VVEFLITVLLOYREFITRPREVNAKSLPLNDEDEEDVAREKORILIDGGQNDLIELKELKI	1920
Qy	1921	YRRKRRKAVDRICVGIPIPGCGFGLCVNGAKSSITPKMLTGDVTYTRGDAFLNRNSILSN	1980
Dp	1921	YRRKRRKAVDRICVGIPIPGCGFGLCVNGAKSSITPKMLTGDVTYTRGDAFLNRNSILSN	1980
Qy	1981	IHEVHOMGKCPQFDATITELTITREHVEFFALLRCVPEKEVKGXGEMAKRLGLGVKXGK	2040
Dp	1981	IHEVHOMGKCPQFDATITELTITREHVEFFALLRCVPEKEVKGXGEMAKRLGLGVKXGK	2040
Qy	2041	YAGNYSGKNRKSLSTAMALIGBPVVEFDEPTTGMDPKARRPLMNCALSVKKEGSRVLT	2100
Dp	2041	YAGNYSGKNRKSLSTAMALIGBPVVEFDEPTTGMDPKARRPLMNCALSVKKEGSRVLT	2100
Qy	2101	SHSHEBEALCTSMALIMVNGRFRCLGSYOHILKNRFGDGYTIVRAGSNPDLAKVYDOPFG	2160
Dp	2101	SHSHEBEALCTSMALIMVNGRFRCLGSYOHILKNRFGDGYTIVRAGSNPDLAKVYDOPFG	2160

Qy	2161	LAFGSGVJLKEKHRRMLQYOLPSSLSSTARFSTLSQSKRRLHEDYSVQOTLLDOPVNF	2220
Db	2161	LAFGSGVJLKEKHRRMLQYOLPSSLSSTARFSTLSQSKRRLHEDYSVQOTLLDOPVNF	2220
Qy	2221	AKDOSDDHLKDLISLHKNGVNVAVLTSFLQDEKVESY	2261
Db	2221	AKDOSDDHLKDLISLHKNGVNVAVLTSFLQDEKVESY	2261

RESULT 13	
AAB38115	
ID	AAB38115 standard; Protein; 2261 AA

DT 29-JAN-2001 (first entry)

Human ABC1 cholesterol transporter mutant, R1587K.

Human ABC1 cholesterol transporter; chromosome 9q31; ATP-binding cassette; HDL deficiency; high density lipoprotein transfer disease; MD; familial HDL deficiency; FHL; polymorphism; cardiovascular disease; coronary artery disease; coronary restenosis; cerebrovascular disease; periparturient vascular disease; Alzheimer's disease; Niemann-Pick disease; Huntington's disease; X-linked adrenoleukodystrophy; cancer; gene therapy; genetic diagnosis; prognosis; prophylaxis; drug screening; transgenic animal; mutant; melanin.

Homo sapiens.

W0200055318-A2.

21-SEP-2000.

15-MAR-2000; 2000WO-IB00532.

15-MAR-1999; 99US-0124702.

17-JUN-1999; 99US-0139600.

(XENO-) XENON BIORESEARCH INC.

Hayden MR, Wilson AR, Pimstone SN;

WPI; 2000-587528/55.

New ABC1 polypeptide is useful for treating diseases associated with

disease and cancer -

Examples; Page -; 229pp; English.

The invention relates to the human ABC1 cholesterol transporter protein (B38082) and to nucleic acid sequences (c69120) which encode it. ABC1 is a member of the ATP-binding cassette (ABC transporter) superfamily of proteins, and plays a crucial role in cholesterol transport, particularly intracellular cholesterol trafficking in monocytes and fibroblasts, being involved in cholesterol efflux from the cell. The gene encoding ABC1 is located on chromosome 9q31, and mutations in this gene are associated with two genetic HDL (high density lipoprotein) deficiency disorders, Tangier disease (TD) and familial HDL deficiency (FHA). These diseases are distinguishable in that TD is an autosomal recessive disorder, while FHA is inherited as an autosomal dominant trait. Low levels of HDL ("good cholesterol") in the blood correlate with a high risk of cardiovascular disease, particularly coronary artery disease, but also cerebrovascular disease, coronary restenosis, and peripheral vascular disease. Conversely, a high level of HDL has protective effects against cardiovascular disease. The invention provides genetic constructs and transgenic cells and non-human animals comprising human ABC1 nucleic

CC acids, and methods of gene therapy for the treatment or prevention of
CC cardiovascular disease comprising the administration of an expression
CC vector encoding ABC1 or an active fragment thereof. The invention also
CC encompasses compounds which mimic ABC1 activity, compounds which
CC stimulate ABC1 expression and methods of screening for such compounds.
CC It further relates to methods for determining whether a patient has an
CC increased risk for cardiovascular disease due to polymorphisms in the
CC ABC1 gene. Human ABC1 proteins and nucleotides can be used to treat
CC or prevent cardiovascular disease, especially coronary artery disease,
CC cerebrovascular disease, coronary stenosis or peripheral vascular
CC disease. They may also be used in the treatment of diseases associated
CC with ABC1 biological activity, such as Alzheimer's disease, Niemann-Pick
CC disease, Huntington's disease, X-linked adrenoleukodystrophy and cancer.
CC The invention specifically excludes proteins with the exact amino acid
CC sequences of GenBank Accession No. CA10005.1 and X75926.1. The
CC acid with the exact sequence as GenBank Accession No. A012376.1. The
CC present sequence represents a mutant human ABC1 cholesterol transporter
CC associated with an altered cholesterol level and therefore an altered
CC risk of cardiovascular disease.
CC Note: The present sequence is not shown in the specification, but is
CC derived from the native human ABC1 shown on pages 152-157.

XX Sequence 2261 AA:

Query Match 99.9%; Score 11786; DB 21; Length 2261;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 2257; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 MACWPOLRLMKNTLERRRQTCQLLEVAMPFLFILISVRLSYRPEDEHCFENKA 60
DB 1 MACWPOLRLMKNTLERRRQTCQLLEVAMPFLFILISVRLSYRPEDEHCFENKA 60
QY 61 MSACPLPWPVGGITCNANPCFRPTPGEPAGVGNFNKSTVARLFSARRLIYSQDQ 120
DB 61 MSACPLPWPVGGITCNANPCFRPTPGEPAGVGNFNKSTVARLFSARRLIYSQDQ 120
QY 121 SMKDRKRYLTLQOIKKSSNLKLODFLVNDFESGLYHNLSPKSTYDKMLRADVILH 180
DB 121 SMKDRKRYLTLQOIKKSSNLKLODFLVNDFESGLYHNLSPKSTYDKMLRADVILH 180
QY 181 KVFLOGYQHLTSLCNGSKSEEMIOLODOVESELCGRKREKLAARLRSMIDLKFTL 240
DB 181 KVFLOGYQHLTSLCNGSKSEEMIOLODOVESELCGRKREKLAARLRSMIDLKFTL 240
QY 241 RLNSTSPPEKELAEATKTLHSGLTLOELFSMRSDMRQEWPLTNVSSSSSTQI 300
DB 241 RLNSTSPPEKELAEATKTLHSGLTLOELFSMRSDMRQEWPLTNVSSSSSTQI 300
QY 301 YQAVSRIVCGHEGGGLIKSLNMYEDNNYKALFGNGTEEDAEFTYDNSTTPYCNDLMK 360
DB 301 YQAVSRIVCGHEGGGLIKSLNMYEDNNYKALFGNGTEEDAEFTYDNSTTPYCNDLMK 360
QY 361 NLESPSLRIKALKPLLVGKILYTPPTPATROYMAEYVNTFOGLAVFHOLEGMEELS 420
DB 361 NLESPSLRIKALKPLLVGKILYTPPTPATROYMAEYVNTFOGLAVFHOLEGMEELS 420
QY 421 PKIWTPEMSQMDLVRLMLDSRDNDHFEQOQDGLDMTAQDIYAFLLKHHPDVSSNGS 480
DB 421 PKIWTPEMSQMDLVRLMLDSRDNDHFEQOQDGLDMTAQDIYAFLLKHHPDVSSNGS 480
QY 481 VYTWREAFNETQOAIKRTISRFMECVNLKLEPIATEVVLINKSMELDERKFMAGIVETG 540
DB 481 VYTWREAFNETQOAIKRTISRFMECVNLKLEPIATEVVLINKSMELDERKFMAGIVETG 540
QY 541 TTPGSTELPHNHKKYKIRMDIDNVERTNKTKDGYWDPGRPADEEDMKRYWGGFATLDQV 600
DB 541 TTPGSTELPHNHKKYKIRMDIDNVERTNKTKDGYWDPGRPADEEDMKRYWGGFATLDQV 600
QY 601 EDAIIRVLGTGKTKGVYQOMPFCYVDDIFLRYMSRSMPLFMTLWYISAVVITIKGI 660
DB 601 EDAIIRVLGTGKTKGVYQOMPFCYVDDIFLRYMSRSMPLFMTLWYISAVVITIKGI 660
QY 661 YEKEARKETMRIMGLDNLIMFWSFISLLPLVLSAGLLVILKGLNLLPYSDRSVVFV 720

DB 661 YEKEARKETMRIMGLDNLIMFWSFISLLPLVLSAGLLVILKGLNLLPYSDRSVVFV 720
QY 721 FLVSFAVYVTLQCFLLISTLSFRANLAACGGITFFLYLPYLVCAWQYVGTLIKTPAS 780
DB 721 FLVSFAVYVTLQCFLLISTLSFRANLAACGGITFFLYLPYLVCAWQYVGTLIKTPAS 780
QY 781 LLSVAFGFCCEFALEEDGIGVQNDNLFESVFEEDGNLTTSVSMLEPDLFGVMTW 840
DB 781 LLSVAFGFCCEFALEEDGIGVQNDNLFESVFEEDGNLTTSVSMLEPDLFGVMTW 840
QY 841 YIAVFPQGYGIDPRWYFPCTKSYWGEESDEKSHDQSNOKRMSICEEMEPHLKLGVS 900
DB 841 YIAVFPQGYGIDPRWYFPCTKSYWGEESDEKSHDQSNOKRMSICEEMEPHLKLGVS 900
QY 901 IONLVKRYRGMKVAVYDGLANFYEQITSFIGNAGKTTMSILTLGLEPPTSGATYLL 960
DB 901 IONLVKRYRGMKVAVYDGLANFYEQITSFIGNAGKTTMSILTLGLEPPTSGATYLL 960
QY 961 GKDIRSEMTIRONLGYCPOHNVLPDLTYEHIHMFYARLKGESEKHVAEMQMALDVG 1020
DB 961 GKDIRSEMTIRONLGYCPOHNVLPDLTYEHIHMFYARLKGESEKHVAEMQMALDVG 1020
QY 1021 LPSSKLKSKTSQLSGGMQRKLSVALAFVGGSKVYLDEPTAGVDPYSRGIWELLKRYQ 1080
DB 1021 LPSSKLKSKTSQLSGGMQRKLSVALAFVGGSKVYLDEPTAGVDPYSRGIWELLKRYQ 1080
QY 1081 GRTITLSTHMDADVGDRAIATISHGKLCCVSSSLFLKNQLOGTGYLLLYKKDVSSLS 1140
DB 1081 GRTITLSTHMDADVGDRAIATISHGKLCCVSSSLFLKNQLOGTGYLLLYKKDVSSLS 1140
QY 1141 SCRRSSSTVSYLKKEDESVSSSDAGISDHESDTLTIDVSAISNLRKHSARLVEDI 1200
DB 1141 SCRRSSSTVSYLKKEDESVSSSDAGISDHESDTLTIDVSAISNLRKHSARLVEDI 1200
QY 1201 GHELTLYLPEAKEGAFVLEFHEIDRLSDLGISSTGSETTLEETFLKVAEESVDAB 1260
DB 1201 GHELTLYLPEAKEGAFVLEFHEIDRLSDLGISSTGSETTLEETFLKVAEESVDAB 1260
QY 1261 TSDGTLPARNRRAFGKQSCLRPFTEDDAADPDSIDDESSETDLISGMDKGSYQAK 1320
DB 1261 TSDGTLPARNRRAFGKQSCLRPFTEDDAADPDSIDDESSETDLISGMDKGSYQAK 1320
QY 1321 GWKLTQOQFVALLMKRLIARRSRKGFPAQIVLPAVFCIALVFSLLVPFGKYPSLEIQ 1380
DB 1321 GWKLTQOQFVALLMKRLIARRSRKGFPAQIVLPAVFCIALVFSLLVPFGKYPSLEIQ 1380
QY 1381 PWTNEDQYTVSNDAPEDTGTLELNLALTKDPGFGTRCMEGNP1PDTPCQAGEEWTTAP 1440
DB 1381 PWTNEDQYTVSNDAPEDTGTLELNLALTKDPGFGTRCMEGNP1PDTPCQAGEEWTTAP 1440
QY 1441 VPOTIMDLFONGMTNMPNSPACOCSSDKIKKMLPYCPACGILPPOROKONTADILLOD 1500
DB 1441 VPOTIMDLFONGMTNMPNSPACOCSSDKIKKMLPYCPACGILPPOROKONTADILLOD 1500
QY 1501 TGRNIDSYLVKTYVOIITAKSLKNTIWNFEFRYGGFSLGVSNTQALPPSOEVNDAIKOMK 1560
DB 1501 TGRNIDSYLVKTYVOIITAKSLKNTIWNFEFRYGGFSLGVSNTQALPPSOEVNDAIKOMK 1560
QY 1561 HUKLAKSSADRELNSLGRFMTGLTTRNNYKWFENKGMHAISSPLNVLNNAILRANLQK 1620
DB 1561 HUKLAKSSADRELNSLGRFMTGLTTRNNYKWFENKGMHAISSPLNVLNNAILRANLQK 1620
QY 1621 GENPSHYGITAENHPLNLTQOULSEVALMTTSYDVLSICVIFRMSVVPASFVFLQER 1680
DB 1621 GENPSHYGITAENHPLNLTQOULSEVALMTTSYDVLSICVIFRMSVVPASFVFLQER 1680
QY 1681 VSKAKHLQFISGVKPVLYWLSNFWDMCNVVPATLVIIIFIGFOOKSYVSTNLPVIAL 1740
DB 1681 VSKAKHLQFISGVKPVLYWLSNFWDMCNVVPATLVIIIFIGFOOKSYVSTNLPVIAL 1740
QY 1741 LLLYGSITPPLAMPASFVKRIPSTAVVLTSVULFINGSAVTYVLELFTONKLANIN 1800


```

Db 1741 LLLXGMSITLMPASTVFKIPSTAYVLTSLVLFITGNSVATFVLELFTIDKLNIN 1800
QY 1801 DILKSVFLIFPHFCLGRGLDMVRNQMADALEFRGGRFVSPLSMDLVGRNLFAMAEG 1860
Db 1801 DILKSVFLIFPHFCLGRGLDMVRNQMADALEFRGGRFVSPLSMDLVGRNLFAMAEG 1860
QY 1861 VVFELITVLYOYRFFIRPRPNNAKLSPLNDEDEVRERORILDDGGONDILEIKELTKI 1920
Db 1861 VVFELITVLYOYRFFIRPRPNNAKLSPLNDEDEVRERORILDDGGONDILEIKELTKI 1920
QY 1921 YRRRRKPAVDRIICGIPGECFGLGNGAGKSTPFKMLTGDITVTGDAFLKNKSTLSN 1980
Db 1921 YRRRRKPAVDRIICGIPGECFGLGNGAGKSTPFKMLTGDITVTGDAFLKNKSTLSN 1980
QY 1981 IHEVHONMGYCPQPDATITELLTGREHVEFPALLRGVPEKEVKGWMAIKRLGLVKYGEK 2040
Db 1981 IHEVHONMGYCPQPDATITELLTGREHVEFPALLRGVPEKEVKGWMAIKRLGLVKYGEK 2040
QY 2041 YAGNVSQGNKKRLSTAMALIGCPVVFLEPPTGMDPKARREFLMNCALSVKKEGRSVLT 2100
Db 2041 YAGNVSQGNKKRLSTAMALIGCPVVFLEPPTGMDPKARREFLMNCALSVKKEGRSVLT 2100
QY 2101 SHSMECEALCTRNAIMVNGFRCLGSVOHLKNRFGDGTIIVRIAGSNPDLKPVDFFG 2160
Db 2101 SHSMECEALCTRNAIMVNGFRCLGSVOHLKNRFGDGTIIVRIAGSNPDLKPVDFFG 2160
QY 2161 LAFPGSVLKEKRRMLOQOLPSSLSLARIFSLISOSKKRLHIEDYSVOTILDQVVFNF 2220
Db 2161 LAFPGSVLKEKRRMLOQOLPSSLSLARIFSLISOSKKRLHIEDYSVOTILDQVVFNF 2220
QY 2221 AKDQSDDDHLKDLSLHKNOYVDVAVLTSFLODEKKESEYV 2261
Db 2221 AKDQSDDDHLKDLSLHKNOYVDVAVLTSFLODEKKESEYV 2261

RESULT 14
AAB38105
ID AAB38105 standard; Protein: 2261 AA.
XX
AC AAB38105;
XX
DT 29-JAN-2001 (first entry)
XX
DE Human ABC1 cholesterol transporter TD-2 mutant protein (Q597R).
XX
KW Human ABC1 cholesterol transporter; chromosome 9q31.
KW ATP-binding cassette, HDL deficiency disorder; high density lipoprotein;
KW Tangier disease; TD; familial HDL deficiency; FHA; polymorphisms;
KW cardiovascular disease; coronary artery disease; coronary restenosis;
KW cerebrovascular disease; peripheral vascular disease;
KW Alzheimer's disease; Niemann-Pick disease; Huntington's disease;
KW X-linked adrenoleukodystrophy; cancer; gene therapy; genetic diagnosis;
KW prognosis; propylaxis; drug screening; transgenic animal; mutant;
KW mutant.
XX
OS Homo sapiens.
XX
PN MO20005318-A2.
XX
PD 21-SEP-2000.
XX
PE 15-MAR-2000; 2000WO-1B00532.
XX
PF 15-MAR-1999; 99US-0124702.
PR 08-JUN-1999; 99US-0138048.
PR 17-JUN-1999; 99US-0139600.
PR 01-SEP-1999; 99US-0151977.
XX
PA (UYBR-) UNIV BRITISH COLUMBIA.
PA (XENO-) XENON BIORESEARCH INC.
XX
PI Hayden MR, Wilson AR, Pimstone SN.
XX

```

```

DR WPI: 2000-587528/55.
DR N-PSDB: AAC69386.
XX
PT New ABC1 polypeptide is useful for treating diseases associated with
PT ABC1 biological activity, e.g. Alzheimer's disease, Huntington's
PT disease and cancer.
XX
PS Examples: Page -: 229pp; English.
XX
CC The invention relates to the human ABC1 cholesterol transporter protein
CC (B3082) and to nucleic acid sequences (C69120) which encode it. ABC1 is
CC a member of the ATP-binding cassette (ABC transporter) superfamily of
CC proteins, and plays a crucial role in cholesterol transport, particularly
CC intracellular cholesterol trafficking in monocytes and fibroblasts, being
CC involved in cholesterol efflux from the cell. The gene encoding ABC1 is
CC located on chromosome 9q31, and mutations in this gene are associated
CC with two genetic HDL (high density lipoprotein) deficiency disorders,
CC Tangier disease (TD) and familial HDL deficiency (FHD). These diseases
CC are distinguishable in that TD is an autosomal recessive disorder, while
CC FHA is inherited as an autosomal dominant trait. Low levels of HDL ("good
CC cholesterol") in the blood correlate with a high risk of cardiovascular
CC disease, particularly coronary artery disease, but also cerebrovascular
CC disease, coronary restenosis, and peripheral vascular disease.
CC Conversely, a high level of HDL has protective effects against
CC cardiovascular disease. The invention provides genetic constructs and
CC transgenic cells and non-human animals comprising human ABC1 nucleic
CC acids, and methods of gene therapy for the treatment or prevention of
CC cardiovascular disease comprising the administration of an expression
CC vector encoding ABC1 or an active fragment thereof. The invention also
CC encompasses compounds which mimic ABC1 activity, compounds which
CC stimulate ABC1 expression and methods of screening for such compounds.
CC It further relates to methods for determining whether a patient has an
CC increased risk for cardiovascular disease due to polymorphisms in the
CC ABC1 gene. Human ABC1 proteins and nucleotides can be used to treat
CC or prevent cardiovascular disease, especially coronary artery disease,
CC cerebrovascular disease, coronary restenosis or peripheral vascular
CC disease. They may also be used in the treatment of diseases associated
CC with ABC1 biological activity, such as Alzheimer's disease, Niemann-Pick
CC disease, Huntington's disease, X-linked adrenoleukodystrophy and cancer.
CC The invention specifically excludes proteins with the exact amino acid
CC sequences of Genbank Accession No: CAA10005.1 and X75926, and the nucleic
CC acid with the exact sequence as Genbank Accession No: A012376.1. The
CC present sequence represents a mutant human ABC1 cholesterol transporter
CC associated with an altered cholesterol level and therefore an altered
CC risk of cardiovascular disease.
CC Note: The present sequence is not shown in the specification, but is
CC derived from the native human ABC1 shown on pages 152-157.
XX
SQ Sequence 2261 AA;
XX
Query Match 99.9%; Score 11785; DB 21; Length 2261;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2257; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
QY 1 MACWPDLRLILKKNLFRRRQTCQLLEAVMPLFIFLILISVRLSYRDEQHECFPKA 60
Db 1 MACWPDLRLILKKNLFRRRQTCQLLEAVMPLFIFLILISVRLSYRDEQHECFPKA 60
QY 61 MPSAGTLPWVGITICNANPCFRYPYTGEGARGVGNGENKKSIVARLFSPARULLYSOKDT 120
Db 61 MPSAGTLPWVGITICNANPCFRYPYTGEGARGVGNGENKKSIVARLFSPARULLYSOKDT 120
QY 121 SMKDKRKVLRITQOIIRKSSSNLKLDDPLVDNETFGSLTYHNSLPKSVVDMRLADVTLH 180
Db 121 SMKDKRKVLRITQOIIRKSSSNLKLDDPLVDNETFGSLTYHNSLPKSVVDMRLADVTLH 180
QY 181 KVFLOGYQLHLTSLGSKSEEMIQLDQEVSELGCLPKREKLAARVLRSNMIDILKPL 240
Db 181 KVFLOGYQLHLTSLGSKSEEMIQLDQEVSELGCLPKREKLAARVLRSNMIDILKPL 240
QY 241 RTLNSTSPFPEKEAEARKITLHSLGTLAQELFSKRSMSDMKQEVMTFTNWNSSSSSTQI 300
Db 241 RTLNSTSPFPEKEAEARKITLHSLGTLAQELFSKRSMSDMKQEVMTFTNWNSSSSSTQI 300

```

QY 301 YQAVSRIVCGHEGGGLKIKSLNMYEDNNYKALFGNGTEEDAETFYDNSTTPYCNDLAK 360
 DB 301 YQAVSRIVCGHEGGGLKIKSLNMYEDNNYKALFGNGTEEDAETFYDNSTTPYCNDLAK 360
 QY 361 NLESPSLRIIMKALKPLLVGKILYPTDPATROYMAEVNKTFOELAVFHDLBEGMWEELS 420
 DB 361 NLESPSLRIIMKALKPLLVGKILYPTDPATROYMAEVNKTFOELAVFHDLBEGMWEELS 420
 QY 421 PKIMFEMNSQEMDLVRMLLDSRDNDHFWEQOGLDMDTAODIYAFIAKHEDEVOSSNGS 480
 DB 421 PKIMFEMNSQEMDLVRMLLDSRDNDHFWEQOGLDMDTAODIYAFIAKHEDEVOSSNGS 480
 QY 481 VYTAREAFNETQOARITISREMECVNLKLEPIATEVNLINKSMELDEREFMAGIVTGG 540
 DB 481 VYTAREAFNETQOARITISREMECVNLKLEPIATEVNLINKSMELDEREFMAGIVTGG 540
 QY 541 ITPGSIELPHHYKIKRMDIDNVERTNKIKDGYWDPGRADPFEDMRYWGGFAYLQDV 600
 DB 541 ITPGSIELPHHYKIKRMDIDNVERTNKIKDGYWDPGRADPFEDMRYWGGFAYLQDV 600
 QY 601 BQAIIRVLGTEKKTGYVMQMPYCYVDDIFLRVMSRSMPLFMTLAVISVAVTIKGI 660
 DB 601 BQAIIRVLGTEKKTGYVMQMPYCYVDDIFLRVMSRSMPLFMTLAVISVAVTIKGI 660
 QY 661 YEKEARKETMRIMGLDINSILMFSWFISSLIPLVSAGLIVILKGLMLFYSDPSVYFV 720
 DB 661 YEKEARKETMRIMGLDINSILMFSWFISSLIPLVSAGLIVILKGLMLFYSDPSVYFV 720
 QY 721 FLVSFAVVTIIQCFLLISTLFSRANLAAACGGIYFTLYPLVLCVAMODYVGTFLKIPAS 780
 DB 721 FLVSFAVVTIIQCFLLISTLFSRANLAAACGGIYFTLYPLVLCVAMODYVGTFLKIPAS 780
 QY 781 LLSPAFGGCGCYFALFEEOGIGVOMDLFESPVEDGFENLTTSIMLPTFTLYGVNTW 840
 DB 781 LLSPAFGGCGCYFALFEEOGIGVOMDLFESPVEDGFENLTTSIMLPTFTLYGVNTW 840
 QY 841 YLEAVFPGQYIPRPWYFPCCTKSYWFGESDEKSHPGSNOKRMSICMEEPETHLKGV 900
 DB 841 YLEAVFPGQYIPRPWYFPCCTKSYWFGESDEKSHPGSNOKRMSICMEEPETHLKGV 900
 QY 901 IONLVKVRDGGKAVAVDGLALNFYEGQITTSFLGHNGAGKTTTMSILTLGFPPTSGTAYIL 960
 DB 901 IONLVKVRDGGKAVAVDGLALNFYEGQITTSFLGHNGAGKTTTMSILTLGFPPTSGTAYIL 960
 QY 961 GKDISEMSTIKONLGVCPQHNVLFDMLLVEEHIWFAKLGSEKHVKAEMEQAALDVG 1020
 DB 961 GKDISEMSTIKONLGVCPQHNVLFDMLLVEEHIWFAKLGSEKHVKAEMEQAALDVG 1020
 QY 1021 LPSKSLKSTQSLSGGMORKLSSVALAFVGGSKVVLIDEPAGVDPYSRGIWELLKTYRQ 1080
 DB 1021 LPSKSLKSTQSLSGGMORKLSSVALAFVGGSKVVLIDEPAGVDPYSRGIWELLKTYRQ 1080
 QY 1081 GRTIILSTHMDADYLDGRIALISHKGLCYGSSLPFLKNOLGTYIYTLVKKDVESSL 1140
 DB 1081 GRTIILSTHMDADYLDGRIALISHKGLCYGSSLPFLKNOLGTYIYTLVKKDVESSL 1140
 QY 1141 SCRNSSSTSYLKKEDSVQSSSDAGLSDHSDTLTIDVSAISMLIKKHSEARLVADI 1200
 DB 1141 SCRNSSSTSYLKKEDSVQSSSDAGLSDHSDTLTIDVSAISMLIKKHSEARLVADI 1200
 QY 1201 GHELTYYVLEPAKEGAFVELFHEIDRLSDLGSSYGISETTELEIFLKVAAEESVD 1260
 DB 1201 GHELTYYVLEPAKEGAFVELFHEIDRLSDLGSSYGISETTELEIFLKVAAEESVD 1260
 QY 1261 TSDGTLPARNRARAFGDKOSCLRPTEDDADPNDSDIDPESRETDLLSGMDGKSYOVK 1320
 DB 1261 TSDGTLPARNRARAFGDKOSCLRPTEDDADPNDSDIDPESRETDLLSGMDGKSYOVK 1320
 QY 1321 GKMLTQOQFVALLMKRLIARSRKGFPAQIYLPAPVFCIALVFSLIYPPGKYPSLEQ 1380
 DB 1321 GKMLTQOQFVALLMKRLIARSRKGFPAQIYLPAPVFCIALVFSLIYPPGKYPSLEQ 1380

QY 1381 PMMYNEQYTFVNSDAPEDGTLELNLALTKDPGFGTRCMEGNPIPDTPQAGEEEMWTAP 1440
 DB 1381 PMMYNEQYTFVNSDAPEDGTLELNLALTKDPGFGTRCMEGNPIPDTPQAGEEEMWTAP 1440
 QY 1441 VPOTIMLPONGWNTQONSPACOCSSDKTKMLPVCPPGAGGLPPPOKONTADILQDL 1500
 DB 1441 VPOTIMLPONGWNTQONSPACOCSSDKTKMLPVCPPGAGGLPPPOKONTADILQDL 1500
 QY 1501 TGRNISDYLKTYVOIILAKSLKNIKWNFEFRYGFSLGVSNTDALPPSOEVNDAIKOMK 1560
 DB 1501 TGRNISDYLKTYVOIILAKSLKNIKWNFEFRYGFSLGVSNTDALPPSOEVNDAIKOMK 1560
 QY 1561 HLKLAKDSSADRLNSLGRMTGIDTRNNYKVMFNKGMHAISPLNVINNALIRANLQK 1620
 DB 1561 HLKLAKDSSADRLNSLGRMTGIDTRNNYKVMFNKGMHAISPLNVINNALIRANLQK 1620
 QY 1621 GENPSHGTAFNHPNLUTQOJLSEVALMTTSDVLVSCVIFAMSFVPASFVFLQER 1680
 DB 1621 GENPSHGTAFNHPNLUTQOJLSEVALMTTSDVLVSCVIFAMSFVPASFVFLQER 1680
 QY 1681 VSKAKHLOFISGVKPYIYMLSNFVMDKCNVBPATLVIIIFICEQOKSYVSSNTLPVL 1740
 DB 1681 VSKAKHLOFISGVKPYIYMLSNFVMDKCNVBPATLVIIIFICEQOKSYVSSNTLPVL 1740
 QY 1741 LLLLYGMSITPLMYPASFVEKIPSTAYVLTSVNLFIGINGSVATEVLELFTDNKLNIN 1800
 DB 1741 LLLLYGMSITPLMYPASFVEKIPSTAYVLTSVNLFIGINGSVATEVLELFTDNKLNIN 1800
 QY 1801 DILKSVFLIFPHCLGGLIDWYKNQAMADALERFGENRFVSPSLMDLVGRNLFAMAVEG 1860
 DB 1801 DILKSVFLIFPHCLGGLIDWYKNQAMADALERFGENRFVSPSLMDLVGRNLFAMAVEG 1860
 QY 1861 VVFLITVLIOYREFIRPRVNAKLSPLNDEDEVDREERORILLDGGQNDILEIKELTKI 1920
 DB 1861 VVFLITVLIOYREFIRPRVNAKLSPLNDEDEVDREERORILLDGGQNDILEIKELTKI 1920
 QY 1921 YRRKRKPAVDRIKCGIPGECFGLGVNGAGKSTFKMLGDTTVPGRDAFLNKNLSILN 1980
 DB 1921 YRRKRKPAVDRIKCGIPGECFGLGVNGAGKSTFKMLGDTTVPGRDAFLNKNLSILN 1980
 QY 1981 IHEYHOMKGYCPQFDATTELLTGREHVEFFALLRGVPEKVGAVGEMATKLLVYGER 2040
 DB 1981 IHEYHOMKGYCPQFDATTELLTGREHVEFFALLRGVPEKVGAVGEMATKLLVYGER 2040
 QY 2041 YACNYSNGNKRKLSTAAALIGCPVYVFLDEPTTGMDPKARFLMNCALSYVKGSRVLT 2100
 DB 2041 YACNYSNGNKRKLSTAAALIGCPVYVFLDEPTTGMDPKARFLMNCALSYVKGSRVLT 2100
 QY 2101 SHSMECEALCTRNAIMVNGRFRCIGSVQHLKKNRFGDYTIIVRIAGSNPDLKPVODPFG 2160
 DB 2101 SHSMECEALCTRNAIMVNGRFRCIGSVQHLKKNRFGDYTIIVRIAGSNPDLKPVODPFG 2160
 QY 2161 LAFPGSVLKEKHRMLOYOLPSSLSLARIPTLSOSKKLHIEDYVSQTTLDQYFVNF 2220
 DB 2161 LAFPGSVLKEKHRMLOYOLPSSLSLARIPTLSOSKKLHIEDYVSQTTLDQYFVNF 2220
 QY 2221 AKDOSDDHLKDLSLHKNOTVVDVAVLTSLFODEKYESVY 2261
 DB 2221 AKDOSDDHLKDLSLHKNOTVVDVAVLTSLFODEKYESVY 2261

RESULT 15
 AAB38110
 ID AAB38110 standard; protein; 2261 AA.
 XX
 AC AAB38110;
 XX
 DT 29-JAN-2001 (first entry)
 XX
 DE Human ABC1 cholesterol transporter mutant; V399A.
 XX
 KW Human ABC1 cholesterol transporter; chromosome 9q31;
 ATP-binding cassette; HDL deficiency disorder; high density lipoprotein;

KW Tangier disease; TD; familial HDL deficiency; FHA; polymorphism;
 KW cardiovascular disease; coronary artery disease; coronary restenosis;
 KW cerebrovascular disease; peripheral vascular disease;
 KW Alzheimer's disease; Niemann-Pick disease; Huntington's disease;
 KW X-linked adrenoleukodystrophy; cancer; gene therapy; genetic diagnosis;
 KW prognosis; prophylaxis; drug screening; transgenic animal; mutant;
 KW
 KW
 OS Homo sapiens.
 PN WO20005318-A2.
 PD 21-SEP-2000.
 PF 15-MAR-2000; 2000WO-IB00532.
 PR 15-MAR-1999; 99US-0124702.
 PR 08-JUN-1999; 99US-0138048.
 PR 17-JUN-1999; 99US-0138600.
 PR 01-SEP-1999; 99US-0151977.
 XX
 PA (UYBR-) UNIV BRITISH COLUMBIA.
 PA (XENO-) XENON BIORESEARCH INC.
 PI Hayden MR, Wilson AR, Plimstone SN;
 XX
 XX WPI: 2000-587528/55.
 PT New ABC1 polypeptide is useful for treating diseases associated with
 PT ABC1 biological activity, e.g. Alzheimer's disease, Huntington's
 PT disease and cancer -
 PS
 PS Examples: Page -: 229p; English.

CC The invention relates to the human ABC1 cholesterol transporter protein
 CC (B36082) and to nucleic acid sequences (C69120) which encode it. ABC1 is
 CC a member of the ATP-binding cassette (ABC transporter) superfamily of
 CC proteins, and plays a crucial role in cholesterol transport, particularly
 CC intracellular cholesterol trafficking in monocytes and fibroblasts, being
 CC involved in cholesterol efflux from the cell. The gene encoding ABC1 is
 CC located on chromosome 9q31, and mutations in this gene are associated
 CC with two genetic HDL (high density lipoprotein) deficiency disorders,
 CC Tangier disease (TD) and familial HDL deficiency (FHD). These diseases
 CC are distinguishable in that TD is an autosomal recessive disorder, while
 CC FHD is inherited as an autosomal dominant trait. Low levels of HDL ("good
 CC cholesterol") in the blood correlate with a high risk of cardiovascular
 CC disease, particularly coronary artery disease, but also cerebrovascular
 CC disease, coronary restenosis and peripheral vascular disease.
 CC Conversely, a high level of HDL has protective effects against
 CC cardiovascular disease. The invention provides genetic constructs and
 CC transgenic cells and non-human animals comprising human ABC1 nucleic
 CC acids, and methods of gene therapy for the treatment or prevention of
 CC cardiovascular disease comprising the administration of an expression
 CC vector encoding ABC1 or an active fragment thereof. The invention also
 CC encompasses compounds which mimic ABC1 activity, compounds which
 CC stimulate ABC1 expression and methods of screening for such compounds.
 CC It further relates to methods for determining whether a patient has an
 CC increased risk for cardiovascular disease due to polymorphisms in the
 CC ABC1 gene. Human ABC1 proteins and nucleotides can be used to treat
 CC or prevent cardiovascular disease, especially coronary artery disease,
 CC cerebrovascular disease, coronary restenosis or peripheral vascular
 CC disease. They may also be used in the treatment of diseases associated
 CC with ABC1 biological activity, such as Alzheimer's disease, Niemann-Pick
 CC disease, Huntington's disease, X-linked adrenoleukodystrophy and cancer.
 CC The invention specifically excludes proteins with the exact amino acid
 CC sequences of Genbank Accession No: CAA10005.1 and X75926, and the nucleic
 CC acid with the exact sequence as Genbank Accession No. A0102376.1. The
 CC present sequence represents a mutant human ABC1 cholesterol transporter
 CC associated with an altered cholesterol level and therefore an altered
 CC risk of cardiovascular disease.
 CC Note: The present sequence is not shown in the specification, but is
 CC derived from the native human ABC1 shown on pages 152-157.
 XX

Seq	Sequence	2261 AA:	99.9%;	Score 11785;	DB 21;	Length 2261;
Query Match						
Best Local Similarity			99.8%;	Pred. No. 0;		
Matches 2257;	Conservative	3;	Mismatches	1;	Indels	0; Gaps 0;
QY	1	MACWPOLRLILMKNLPRRRRCQCLLEAVMPLFTFLIISVSLPPVQEGHCHPNKA	60			
DB	1	MACWPOLRLILMKNLPRRRRCQCLLEAVMPLFTFLIISVSLPPVQEGHCHPNKA	60			
QY	61	MPASGTLPPWOGIICNANNPCFRYPPEAGVGVGNENKSIYARLPSDARLLVLSQKPT	120			
DB	61	MPASGTLPPWOGIICNANNPCFRYPPEAGVGVGNENKSIYARLPSDARLLVLSQKPT	120			
QY	121	SMKDMKRVLTLOQIKSSSNLKLQDFLVNDFPFGFLYHNLSLPKSTYDKMLRADYILH	180			
DB	121	SMKDMKRVLTLOQIKSSSNLKLQDFLVNDFPFGFLYHNLSLPKSTYDKMLRADYILH	180			
QY	181	KVFLQGYQLHTSLCNKSKSEEMTOLGDOVESELGSLPKREKLAARVYLRSMNDILKPTL	240			
DB	181	KVFLQGYQLHTSLCNKSKSEEMTOLGDOVESELGSLPKREKLAARVYLRSMNDILKPTL	240			
QY	241	RTLNSTSPFPSKELAEATKTLHSLGTLAQLFMSMSMSDMRQEVMLTNVNSSSTQI	300			
DB	241	RTLNSTSPFPSKELAEATKTLHSLGTLAQLFMSMSMSDMRQEVMLTNVNSSSTQI	300			
QY	301	YQANSRTVCGHPGCGGLIKSLMWYEDNNYKALFGNGTGEDEAFETPDNSTTPYCDMLK	360			
DB	301	YQANSRTVCGHPGCGGLIKSLMWYEDNNYKALFGNGTGEDEAFETPDNSTTPYCDMLK	360			
QY	361	NLSSPSLRITMKALKPLVGLKILYTPDPTRQVAEVAVKTPQELAVNHDLEGMEELS	420			
DB	361	NLSSPSLRITMKALKPLVGLKILYTPDPTRQVAEVAVKTPQELAVNHDLEGMEELS	420			
QY	421	PKIWTPEMSQENDVLMLLDSRDNDHFWEQDLGIDMTADQIVAFIAHPEDVOSSNGS	480			
DB	421	PKIWTPEMSQENDVLMLLDSRDNDHFWEQDLGIDMTADQIVAFIAHPEDVOSSNGS	480			
QY	481	VYTMREAFNETNOAIRISFMFECVNLKDEPATEVWILNKSMELLDRKKFMAGIVFG	540			
DB	481	VYTMREAFNETNOAIRISFMFECVNLKDEPATEVWILNKSMELLDRKKFMAGIVFG	540			
QY	541	ITPGSIELPHVKKYKIRMDIDNERTNKIKDGYWDGPRADPEDMRYVWGGFAYLDQV	600			
DB	541	ITPGSIELPHVKKYKIRMDIDNERTNKIKDGYWDGPRADPEDMRYVWGGFAYLDQV	600			
QY	601	EQAIRVLTTEKKTGYVQMOQMPYPCYVDIFLRVMSRMPLEMTLAWITYSAVVIKGI	660			
DB	601	EQAIRVLTTEKKTGYVQMOQMPYPCYVDIFLRVMSRMPLEMTLAWITYSAVVIKGI	660			
QY	661	YKEARLKEFMRIIMGDINSILMFSWFISSILPLVSAGLVITLKGNTLPYSDPSVVF	720			
DB	661	YKEARLKEFMRIIMGDINSILMFSWFISSILPLVSAGLVITLKGNTLPYSDPSVVF	720			
QY	721	FLSVFAVVTILQCLFISTLFSRANLAAACGGIYFTFLVPLVCVAMQDYVGFYTLKIFAS	780			
DB	721	FLSVFAVVTILQCLFISTLFSRANLAAACGGIYFTFLVPLVCVAMQDYVGFYTLKIFAS	780			
QY	781	LISVPAFGCECFALPEEBOGTGVOMDNLFESPVDEEDGNLTSTSMMLFDPIFVGMVW	840			
DB	781	LISVPAFGCECFALPEEBOGTGVOMDNLFESPVDEEDGNLTSTSMMLFDPIFVGMVW	840			
QY	841	YIYAVPFGQYGIIRPWFYFCTKSYWFGESDEKSHFGSNQKRMSEICMEDEPTHLKIGVS	900			
DB	841	YIYAVPFGQYGIIRPWFYFCTKSYWFGESDEKSHFGSNQKRMSEICMEDEPTHLKIGVS	900			
QY	901	IQMLVAVYRDGMKVAVDGALANFYEGQITSPFGHNGAGKTTTMSITLGLPPTSGTAYIL	960			
DB	901	IQMLVAVYRDGMKVAVDGALANFYEGQITSPFGHNGAGKTTTMSITLGLPPTSGTAYIL	960			
QY	961	GKDIRSEMSITRONLGVCPQHNVLPDMLVVEEHIMVYARLKGISEKHVAEKEQNALDVG	1020			
DB	961	GKDIRSEMSITRONLGVCPQHNVLPDMLVVEEHIMVYARLKGISEKHVAEKEQNALDVG	1020			

```

QY 1021 LPSSKSKTSQSOLSGMOKRLSVALAFVGSKVYILDEPTAGVDPYSKRCITWELLKYRQ 1080
    |||
Db 1021 LPSSKSKTSQSOLSGMOKRLSVALAFVGSKVYILDEPTAGVDPYSKRCITWELLKYRQ 1080
QY 1081 GRTIILSTHMHDEADVLDGRIAIISHGKLCVGSLSFLKNOLGTYVLTLYKKDVESLS 1140
    |||
Db 1081 GRTIILSTHMHDEADVLDGRIAIISHGKLCVGSLSFLKNOLGTYVLTLYKKDVESLS 1140
QY 1141 SCRNSSSTVYLKREDYSOSSDAGLSDBHESDTLTIDVAISINLRKHVSEARLYEDI 1200
    |||
Db 1141 SCRNSSSTVYLKREDYSOSSDAGLSDBHESDTLTIDVAISINLRKHVSEARLYEDI 1200
QY 1201 GHELTYYVPEYAKGAVELFHEIDRSLDGISSYGISSTTELEIFLKYAESGVDAE 1260
    |||
Db 1201 GHELTYYVPEYAKGAVELFHEIDRSLDGISSYGISSTTELEIFLKYAESGVDAE 1260
QY 1261 TSDGTLPARNRNRAFGDKOSCLRPTEDDAADPNDSIDPESRETDLSGMDGKGYQVK 1320
    |||
Db 1261 TSDGTLPARNRNRAFGDKOSCLRPTEDDAADPNDSIDPESRETDLSGMDGKGYQVK 1320
QY 1321 GWKLTQOQFVALLMKRLLIARSRKGFQIYVPAVFCIALVFSLYVPPGKTPSLEQ 1380
    |||
Db 1321 GWKLTQOQFVALLMKRLLIARSRKGFQIYVPAVFCIALVFSLYVPPGKTPSLEQ 1380
QY 1381 PMYNEOYTFVSNDAPEDTGTLELNLATKDPGFGTRCMEGNPIDTPCOAGEEWTAP 1440
    |||
Db 1381 PMYNEOYTFVSNDAPEDTGTLELNLATKDPGFGTRCMEGNPIDTPCOAGEEWTAP 1440
QY 1441 VPQITMDLFQNGNMTMNPSPACQSSDKIKMLPVCPPGAGLPPPORKONTADILQDL 1500
    |||
Db 1441 VPQITMDLFQNGNMTMNPSPACQSSDKIKMLPVCPPGAGLPPPORKONTADILQDL 1500
QY 1501 TGRNISTDIYKTYVOITIAKSLKNKTWNEPFRYGFSLGVSNTOALPSEQEVNDAIKQMK 1560
    |||
Db 1501 TGRNISTDIYKTYVOITIAKSLKNKTWNEPFRYGFSLGVSNTOALPSEQEVNDAIKQMK 1560
QY 1561 HLKAKDSADRFNLNLSGFMGTGDTNRNNVKWFMNKKCMHAISFLANTNATLRANLQK 1620
    |||
Db 1561 HLKAKDSADRFNLNLSGFMGTGDTNRNNVKWFMNKKCMHAISFLANTNATLRANLQK 1620
QY 1621 GENPSHYGITAFNHLNLTQOOLSEVALMTTSVDLVASICVIFAMSFPASFPVFLIOER 1680
    |||
Db 1621 GENPSHYGITAFNHLNLTQOOLSEVALMTTSVDLVASICVIFAMSFPASFPVFLIOER 1680
QY 1681 VSKAKHLQFISGVKPYIYWLNFVMDKCYVPATLVIIIFICFOOKSYVSTNLPVLAL 1740
    |||
Db 1681 VSKAKHLQFISGVKPYIYWLNFVMDKCYVPATLVIIIFICFOOKSYVSTNLPVLAL 1740
QY 1741 LLLLGWSTTPLYMYPASFPVFKIPSTAYVVLTVSNLFIGINGVATFVLELFTDNKLNIN 1800
    |||
Db 1741 LLLLGWSTTPLYMYPASFPVFKIPSTAYVVLTVSNLFIGINGVATFVLELFTDNKLNIN 1800
QY 1801 DILKSVLELFPHFCLGRLIDVKNQAMADALEREGENRFVPLSMDLVGRNLFAMAVEG 1860
    |||
Db 1801 DILKSVLELFPHFCLGRLIDVKNQAMADALEREGENRFVPLSMDLVGRNLFAMAVEG 1860
QY 1861 VVEFLITVLIQYRFELRPPVNAKLSPLNDEDEVDYRERORITLDGGGONDILEIKELTKI 1920
    |||
Db 1861 VVEFLITVLIQYRFELRPPVNAKLSPLNDEDEVDYRERORITLDGGGONDILEIKELTKI 1920
QY 1921 YRRKRPADVRIICVGIPECECGLLGVNGAGKSTFKMLTGDTVTYRGDAFLNKNSILSN 1980
    |||
Db 1921 YRRKRPADVRIICVGIPECECGLLGVNGAGKSTFKMLTGDTVTYRGDAFLNKNSILSN 1980
QY 1981 IHEVHONMGYCPQDAITELLGREGHEFFALLRGVPEKEVGKVBEMAIKRLGLVYKGEK 2040
    |||
Db 1981 IHEVHONMGYCPQDAITELLGREGHEFFALLRGVPEKEVGKVBEMAIKRLGLVYKGEK 2040
QY 2041 YAGNTSGGKKRLSTAMALIGPVPVFLDEPTTGMDPKARFLMNCALSYKESGRSVLT 2100
    |||
Db 2041 YAGNTSGGKKRLSTAMALIGPVPVFLDEPTTGMDPKARFLMNCALSYKESGRSVLT 2100

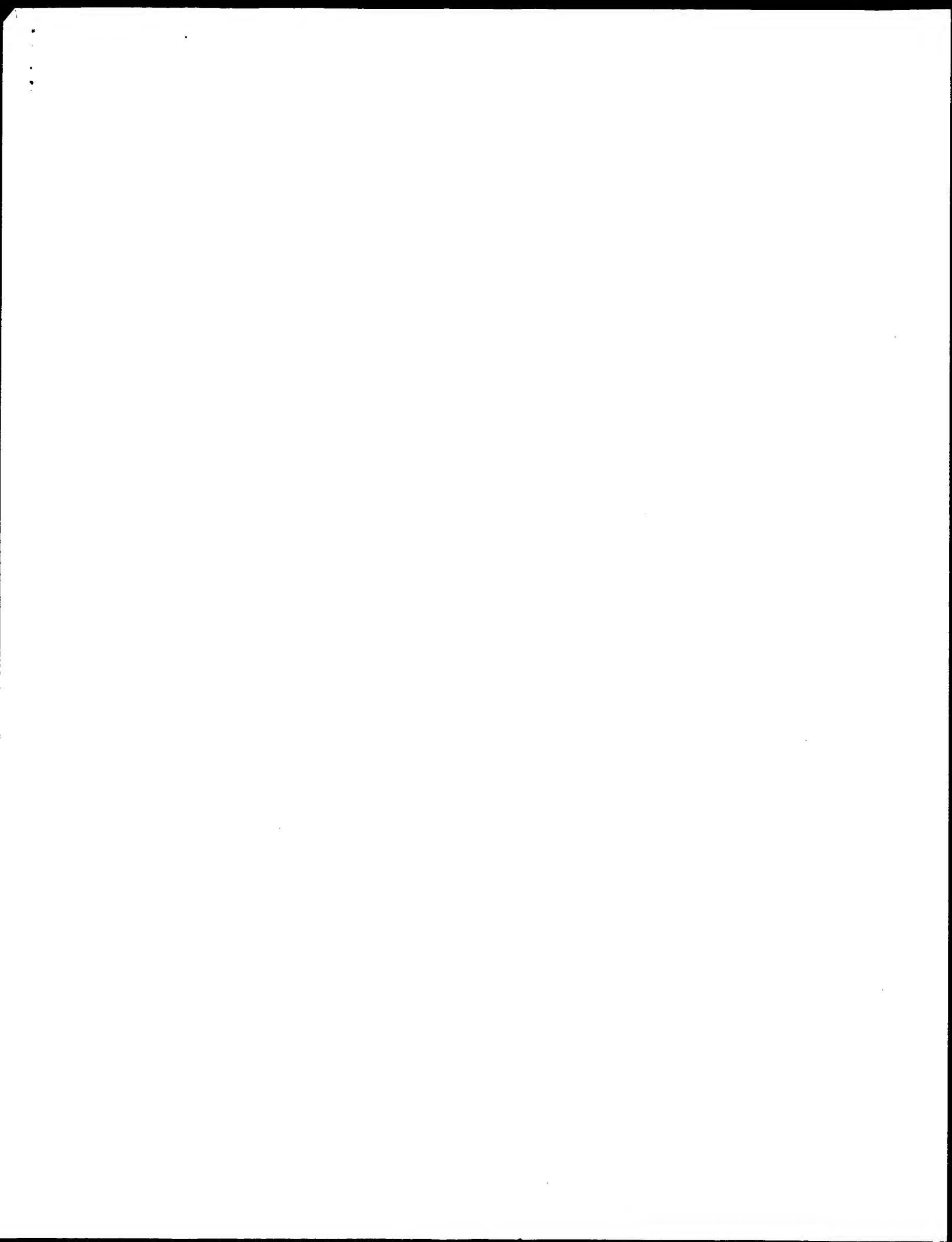
```

```

QY 2101 SHSMECEALCTEMAIWVNGRFRCLGSVOHLKNRFGDGYTIWRIAGSNPDLKPVQDFFG 2160
    |||
Db 2101 SHSMECEALCTEMAIWVNGRFRCLGSVOHLKNRFGDGYTIWRIAGSNPDLKPVQDFFG 2160
QY 2161 LAFPGSVLKEKHHNMLOYLPSLSLARIFFSLSOSKKRLHIEDYSVQOTLDOYFVNF 2220
    |||
Db 2161 LAFPGSVLKEKHHNMLOYLPSLSLARIFFSLSOSKKRLHIEDYSVQOTLDOYFVNF 2220
QY 2221 AKQOSDDHLKDLSLHKNOTVVDVAVUTSFLODEKYESYV 2261
    |||
Db 2221 AKQOSDDHLKDLSLHKNOTVVDVAVUTSFLODEKYESYV 2261

```

Search completed: March 7, 2003, 08:47:08
Job time : 97 secs



Db 1 MAFWTOGLILMKNTYRRROTFOILLIEVAMPLEFFILISVRLSYPRYEOHECFPNKA 60
QY 61 MSAGTLPWVOGICIANNPCCRRYPPGAPAVGVGNFKSIVARLFSPARLLLYSOKDT 120
Db 61 MSAGTLPWVOGICIANNPCCRRYPPGAPAVGVGNFKSIVARLFSPARLLLYSOKDT 120
QY 121 SMKDMKRYARTLOQIKKSSS-NIKLODFVNDNETPSGFYHNLSPKSVOKMIRADYTL 179
Db 121 SIKDVOKVLAKIRKIGNSSGIDLKIRDFLVNDNETPSDLRAHVSNPSSAWEELDAEYVL 180
QY 180 HAVYUOGLHLSICNOSKSEEMTOLGDOY---SE---LGLPEKELAAAEVRLSNMD 234
Db 181 QYVIVSGRIQIRDCNSAISSEFLTIONRSVAMDESAFLCTLPKETHAELAFRANLN 240
QY 235 LKPLIRTNSTSPSPKELAEATKTLHSLGTLAQOELFSRNSRSDMOEWFNLNVSS 294
Db 241 PLKPLQRELEFFNSL--KDLSTVIALDLSGLKVLKELLSKMSWSDMOEWFNLNVAS 298
QY 295 SSSQIYQAVSRIVCGHPEGGLKIKSLNWEEDNNYKALFGNGTEDEDAFTFYDNSTPY 354
Db 299 NSSQIYQAVSRIVCGHPEGGLKIKSLNWEEDNNYKALFGNGTEDEDAFTFYDNSTPY 358
QY 335 CNDLKNLSESPSLRITIKALKPLLVGLITLTPDTPATROYMAEYNTFOELAVFHDEG 414
Db 339 CNELKNLSESPSLRITIKALKPLLVGLITLTPDTPATRIKIMAEVNTFOELAVFHDEG 418
QY 415 MWEELSPKIMTFEMNSQMDLVRLMLDSRDNDHFEMOQDGLMDTADODIVAFIAKHEDV 474
Db 419 MWEELSPKIMTFEMNSQMDLVRLMLDSRDNDHFEMOQDGLMDTADODIVAFIAKHEDV 475
QY 475 OSSNSVYTWAEATNETOALITRISREMECVNLKLEPIATEVWLJNSMELDERKFWA 534
Db 476 EADNOMVYTWAEATNETOALITRISREMECVNLKLEPIATEVWLJNSMELDERKFWA 535
QY 535 GIVEFGIPIGSLTELPHHKXKIRMDINDVERNKIKDGYMDPGRAPPEDEMRKRVWGGA 594
Db 536 GAVEFELIAPNSTELPQHKYKIRMDINDVERNKIKDGYMDPGRAPPEDEMRKRVWGGA 595
QY 595 YLODVVEOAIIRVLGTEKTKGYVMQMPYPCYVDIFLRLRMSRMDLFTMLAVISVAV 654
Db 596 YLODVVEOAIIRVLGTEKTKGYVMQMPYPCYVDIFLRLRMSRMDLFTMLAVISVAV 655
QY 655 IIKGIVYERKARKETMRIMGLDNLIMPSWFISSLLPILVSGLLVTLKLGNLPSYD 714
Db 656 IIKGIVYERKARKETMRIMGLDNLIMPSWFISSLLPILVSGLLVTLKLGNLPSYD 715
QY 715 PSVVEFLSVFAVYTLQCEFLISPLFSRANILAAACGIIYFTLYLPVLCVAMODVYGF 774
Db 716 PSVVEFLISPLISPLISPLISPLISPLISPLISPLISPLISPLISPLISPLISPLIS 775
QY 775 LKIFASLLSPVAFSGCEYFALPEEOGIGVQMDNLFESPEDEGFNLJTSIMLFDJFL 834
Db 776 LKIFASLLSPVAFSGCEYFALPEEOGIGVQMDNLFESPEDEGFNLJTSIMLFDJFL 835
QY 835 YGVATWITIEVFPQOYGIPIPPWYFPCPKSVYGEESPEKSHPSNOKRMEICMEDEPH 894
Db 836 YGVATWITIEVFPQOYGIPIPPWYFPCPKSVYGEESPEKSHPSNOKRMEICMEDEPH 895
QY 895 LKIGVSIQNLVKKYRDMKVAVDGLALNFYEQOITSEFLGNAGAKTJYMSIILGLPPTS 954
Db 896 LKIGVSIQNLVKKYRDMKVAVDGLALNFYEQOITSEFLGNAGAKTJYMSIILGLPPTS 955
QY 955 GTAVIILGKOLRSKMSITRMLGYCPOHNVLEMLJVEBHITWYARLKGISEKHVAEMO 1014
Db 956 GTAVIILGKOLRSKMSITRMLGYCPOHNVLEMLJVEBHITWYARLKGISEKHVAEMO 1015
QY 1015 MALDVLGSSKLSKTSQSGMQRKLSVALAFVSGSKVYIIDEPTAGVDPYSKRIMEL 1074
Db 1016 MALDVLGSSKLSKTSQSGMQRKLSVALAFVSGSKVYIIDEPTAGVDPYSKRIMEL 1074
QY 1075 LKRYROGRTIILSTHMDADVGLDRIALISHGKLCQVSSJFLKNQLOTGYTLVKKD 1134
Db 1075 LKRYROGRTIILSTHMDADVGLDRIALISHGKLCQVSSJFLKNQLOTGYTLVKKD 1134

QY 1135 VESSLSSCRNSSVSYLKKEDSVSOSSDAGLGSHPESDTLTDVATSINLIRKHVSEA 1194
Db 1135 VESSLSSCRNSSVSYLKKEDSVSOSSDAGLGSHPESDTLTDVATSINLIRKHVSEA 1194
QY 1195 RLVEDICHELTYLIPYEAKEGAFVELFHEIDRLSDLGISYGISSETTLEEFILKVAE 1254
Db 1195 RLVEDICHELTYLIPYEAKEGAFVELFHEIDRLSDLGISYGISSETTLEEFILKVAE 1254
QY 1255 SGVDAFTSDCTLPAKRRRRRFGDQOSCLREPTDDADAPDSDIDPESSETLLSSMDK 1314
Db 1255 SGVDAFTSDCTLPAKRRRRRFGDQOSCLREPTDDADAPDSDIDPESSETLLSSMDK 1314
QY 1315 GSIVOKGWLKQOQFALMLKRLLIARSRKGFPAQIVLPAVVCIALVFSLLVPPFGKY 1374
Db 1315 GSIVOKGWLKQOQFALMLKRLLIARSRKGFPAQIVLPAVVCIALVFSLLVPPFGKY 1374
QY 1375 PSLELOPMMTNEQYTVNSNAPEDGTLELNAITKDPQGTROMGNDIPDPCOAGEE 1434
Db 1375 PSLELOPMMTNEQYTVNSNAPEDGTLELNAITKDPQGTROMGNDIPDPCOAGEE 1434
QY 1435 EMTTAPVPTIMDLFONGMNTMONPSPACOSSDKIKMLPVCOPGAGGLPPOROKNTA 1494
Db 1435 EMTTAPVPTIMDLFONGMNTMONPSPACOSSDKIKMLPVCOPGAGGLPPOROKNTA 1494
QY 1495 DIIQDLTGRNISDYLVKTYVOIIAKSLKNNIWNBERYGGFSLGVSNQALPPSOEYND 1554
Db 1494 DIIQDLTGRNISDYLVKTYVOIIAKSLKNNIWNBERYGGFSLGVSNQALPPSOEYND 1553
QY 1555 IKQMKHLKIAKSSADRLNLSIGRMGLDTRNNVKKVFPNNKGMHAISSPLNTVNNAL 1614
Db 1554 IKQMKHLKIAKSSADRLNLSIGRMGLDTRNNVKKVFPNNKGMHAISSPLNTVNNAL 1613
QY 1615 RANLOGENPSSHGIIATFNHPLNLTQOQSEVALMTTSVDVLSICVIFAMFVASFVY 1674
Db 1614 RANLOGENPSSHGIIATFNHPLNLTQOQSEVALMTTSVDVLSICVIFAMFVASFVY 1673
QY 1675 FLIOERYSKAKHLQFISGKRYTYMLSNFWMCMYVVPATLVIIITFCQOQSVYSTN 1734
Db 1674 FLIOERYSKAKHLQFISGKRYTYMLSNFWMCMYVVPATLVIIITFCQOQSVYSTN 1733
QY 1735 LPLVALLLLLGWSITPLMYPASFEFKIPSTAYVVLTVSNLFIGINGSVAFVLEFTDN 1794
Db 1734 LPLVALLLLLGWSITPLMYPASFEFKIPSTAYVVLTVSNLFIGINGSVAFVLEFTDN 1793
QY 1795 KLNINDILKSVFLIPHFCGLGLIDVKKQAMADALERRGENFVSPLSMDLVGRNLF 1854
Db 1794 KLNINDILKSVFLIPHFCGLGLIDVKKQAMADALERRGENFVSPLSMDLVGRNLF 1853
QY 1855 AAVAGVVEFLITVLIQYRFFIRPPRYNAKLSPLNDEDEDEVRREQRILDDGGONDLEI 1914
Db 1854 AAVAGVVEFLITVLIQYRFFIRPPRYNAKLSPLNDEDEDEVRREQRILDDGGONDLEI 1913
QY 1915 KELTIYRRKRPAYDRICVIGIPGECFGLGVNAGAKSSFFKMLTGTDTYTRDAPLNF 1974
Db 1914 KELTIYRRKRPAYDRICVIGIPGECFGLGVNAGAKSSFFKMLTGTDTYTRDAPLNF 1973
QY 1975 NSILSNIEHVQNMKYCQOFDAITELTGRBHVFEFALLKRGVPEKVGKWEAIRRLGL 2034
Db 1974 NSILSNIEHVQNMKYCQOFDAITELTGRBHVFEFALLKRGVPEKVGKWEAIRRLGL 2033
QY 2035 VYKGEKYGANTSGKNRKLSTAMALIGPPVYFLDEPTGMDPKARFPLNMCALSVYKEG 2094
Db 2034 VYKGEKYGANTSGKNRKLSTAMALIGPPVYFLDEPTGMDPKARFPLNMCALSVYKEG 2093
QY 2095 RSVVLTISHMECEALCTRMAIIVNGRRCLGSVOHLKNRFGDGYTLIVRILAGSNPLDKP 2154
Db 2094 RSVVLTISHMECEALCTRMAIIVNGRRCLGSVOHLKNRFGDGYTLIVRILAGSNPLDKP 2153
QY 2155 VODFFGLAFPGSVYLKRRNNLOVQOLPSSLSLARITSIISSQKRLHIEDYVSGQTTLD 2214
Db 2154 VODFFGLAFPGSVYLKRRNNLOVQOLPSSLSLARITSIISSQKRLHIEDYVSGQTTLD 2213

QY 2215 QVFNFAKDDDDHLKDLKLNKQTVVAVLSTLQDEKESYV 2261
 |||||
 Db 2214 QVFNFAKDDDDHLKDLKLNKQTVVAVLSTLQDEKESYV 2260

RESULT 2

035600 PRELIMINARY: PRT: 2310 AA.

AC 035600: AC 035600: ID 035600: AC 035600: DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE ATP-binding cassette transporter.
 GN ABCA4 OR ABCR.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6;
 RX MEDLINE=9745663; PubMed=9202155;
 RA Azarian S.M., Travis G.H.;
 RT "The photoreceptor rim protein is an ABC transporter encoded by the
 RT gene for recessive Stargardt's disease (ABCR).";
 RL FEBS Lett. 409:247-252(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6;
 RA Azarian S.M., Travis G.H.;
 RL Submitted (JUN-1998) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF000149; AAC23916.1; -;
 DR MGD: MGI:109424; Abca4.
 DR InterPro: IPR003593; AAA_ATPase.
 DR InterPro: IPR003439; ABC_transporter.
 DR Pfam: PF00005; ABC_tran; 2.
 DR ProDom: PD000006; ABC_transporter; 2.
 DR SMART: SM00382; AAA; 1.
 DR TIGRFAMs: TIGR01257; rim_protein; 1.
 DR ProSite: PS00211; ABC_TRANSPORTER; UNKNOWN_1.
 KW ATP-binding.
 SQ SEQUENCE 2310 AA; 260207 MW; 8370C6C8A62EF294 CRC64;

Query Match 49.9%; Score 5888; DB 11; Length 2310;
 Best local similarity 50.1%; Pred. No. 0;
 Matches 1168; Conservative 362; Mismatches 624; Indels 178; Gaps 27;

QY 6 QRLLLMKLTLFRRQTCQLLEVAAPLFLILISLVLSYPPYEQHECHFPNKAMPASG 65
 |||||
 Db 6 QIOLLLMKWTLTKRKOKIRFVVELVPLFLYLILRLANANPLYSQHECHFPNKAMPASG 65
 QY 66 TLFWGIIICNNANPCFRPTPEAGRGVGNPKSIVARLFAFSPARLILLYSKQDTSMKM 125
 |||||
 Db 66 LFLWLDGIFCNMNNPCFQNTPEGSPGTVSNYNSTILARVRFQELFMDTPEVHLLGOV 125
 QY 126 RYVLRFLQOI-----KSSSNLKLQDPLVDNETFSGFLYHNLSPKSTVDMKLRAD 176
 |||||
 Db 126 WALRLTLSPQMDLRTNHPREFAGRGLOIRDLKDEALFLFLMRNIGLSDVAHLVNSQ 185
 QY 177 VILHVKVLOGYQLHTSL-CNGSKSEMT---QLGDQEVSE-LGQLPEKTLAAERVL 230
 |||||
 Db 177 VVEQFAVGVPELDELIDICSEALLQRFITFSQRGAQTVRDALCLPSVTLQMTEDTLY 245
 QY 231 SNNDIILKPLRTINSTSPFSKELATKTLHSLGT---AOELFSMRSW-----SDMR- 282
 |||||
 Db 246 ADVDFPK-----LFLVLPFLLDSSSGGTLNLRWGGIISDLSP 282
 QY 283 -----QEVNF-----LTVNNSSSSTQYQAVSRVYCGHPGCGGLIKLSLWYED 327
 |||||
 Db 283 RMQKFIHRSVQDLWVSRLLONGRPETFTQMLSTLSDLCGYPREGGSRVSEFWYED 342
 QY 328 NNTKALFGNGTEDEATFYDNSTPYCNDLKNLLESSPLSRILKALKPLVLGKLLYTP 387
 |||||

Db 343 NNTKALFGIDSTRKPAYSYDKRTTSPCNLSIQLESNPLTKIAMPAAKPLMGKLLYTP 402
 QY 388 DTPATROYMAEVNTPQDLAVFHDLEGMWELSPKIMTFEMENSOEDLVNMLDSDNDH 447
 |||||
 Db 403 DSPARRIMKMANSTFEEDLVKRLVAKAMEVGFQWTFEKSQMTVINDTLOHTPYAD 462
 QY 448 FWEQOLDGLMDTAODIAFLAKHPEDEVSSNGSVYVREAFAENETNOARTISRFMCVNL 507
 |||||
 Db 463 FINQLQEBEGITTEAVLNFPNSNGPQEKQADDMTSFQWRDIFNTDRFLRANQYLEGLVL 522
 QY 508 NKEPIATEWLKSMELLDERKFAAGIYFTGTPGISTELPHVKKIKIMDIDNERTN 567
 |||||
 Db 523 DKFESYDEVDLQTRALSLLEENRFMAGVYFPGYPAWASSLPVHYKKIKIMDIDVEKTN 562
 QY 568 KIKGYMDPGPRADPFEDMRVWGGFAVLQDVYQALIRVLGTGEEKTYGYMOOMPXY 627
 |||||
 Db 583 KIKRWYDSCGRAPVEDFRTIMGFAVLQDMYQGLVKSQMQAEPPIGYLQOMPPCR 642
 QY 628 VDDIFLRVSRSMPLFMTLAWISAVVINGIYERBARLEKTRIMGLDNLSTLWFSMT 687
 |||||
 Db 643 VDDSEMIILNCRPIFLFWLAWISVSMYKGVLEKELRLKETLKNQGSNAVIMCTWFL 702
 QY 688 SSLPLVLSAGLLVYLKGLNLPYSDPSVYFVLSFPAVYTLIOCLTILSTRANLAA 747
 |||||
 Db 703 DSFSIMALSTLPLFLIMHGRILHYSDPFLFLFLAFATATIMQSFLLSTLFSKASLAA 762
 QY 748 ACGGTYFTLPLVYLCAVADYVGFLLKIFASLSLSPVAFGECEYFALFEEDGIGVMD 807
 |||||
 Db 763 ACSGYITFLPLPVLCFAMODRMTADLKTIVSLSSVAFGEFGEYLVREDEGLGQNS 822
 QY 808 NLFPSPVEEDFNLTLSIMLFPDLYGVMTWYIEAVFPQYQIPPMYFPCTKSYWFG 867
 |||||
 Db 823 NIGKSPLEGDFSLSKMKMLDLAALYGLAWLDQVPEDYGTPLPWYFLLOESYWL 882
 QY 868 -----EESDEKSHPGSNQKRAMSETICMEERTHLKLGVSQNLQVY 908
 |||||
 Db 883 GEGCSTRERALEKTEPLTEMEDEPH---EGMDNSFERELPGLVPGVCYKMLKVF 938
 QY 909 RDGMKVAVDGALNPEEGQITSPFGHNGAGKTMTSTLTGTFPTSTAYILGKDISEM 968
 |||||
 Db 939 EPSGRPAVDRLNITFYEQMTAFLGHNGAGKTTLTITGLTPPTSGTVLIGKDIETNL 998
 QY 969 STIRONLQVCPQHNVLFDMLTVEEHIMEYARLKLSEKHYAEMQALDVLGPPSSKLKS 1028
 |||||
 Db 999 DVNRQSLCMPCQHNILFHLTLVAEHLLEYAOLKGRSWEAOLEMEAMLEDTGL-HIKRNE 1057
 QY 1029 KTSQISGMQKRLSALALAFVSGKVIILDEPTAGVDPRSRGIVELLKRYQGTITLST 1088
 |||||
 Db 1058 EAQDLSGDMQKRLSVALAFVDSKVVVLDEPTSGVDPRSRISWDLLKRYSGRTITMT 1117
 QY 1089 HHMDADVLGRIALISHGKLCVSSSLFNKQGLGTGYTLVK--KDES-----SL 1140
 |||||
 Db 1118 HHMDADLIGRIALITSGRILYCSGTPLFLKNCFCGTGYTLVKKMNINQSGRGCGVC 1177
 QY 1141 SCRNSSSTVSYLKKEDSVSQSSDAGLSDHSDTLTIVSAISNLIRKHSEARLVEDI 1200
 |||||
 Db 1178 SCTSKFSFTRCPTRVDEIT-----EEQVLDGDVQELMDLYVHHVPEAKLEECI 1225
 QY 1201 GHELTLYVLYPEAKAGAVELFHEIDRLSDLGISYGISSTIETLEFLKAYAESGVDAE 1260
 |||||
 Db 1226 GOELFLLPNKKKFORAVALSFRELETLADLGSSFCISDTPLEETFLKTYDAGAGSM 1285
 QY 1261 TSDGTLPARRNRAFGDQSC-----LRPTE-----DAAAPPNDSIDIPESERDIL 1308
 |||||
 Db 1266 FYGG---AQQRKEQAGLRHPCSAPTREKLRQYAOARHTCSPPQVDRPKQSPREED 1338
 QY 1309 SGMGKGSYQYKMWLTOQOQVALLMKRLLIARNSRKGFAQIYLPVAVFCIALVPSLIV 1368
 |||||
 Db 1339 -----GVPEFTGARLLIQLHQAALLVYKRFNHTIRSRKDFVAQIYLPVAVFCIALVPSLIV 1393
 QY 1369 PPGKYVPSLELDQPMYNYQYTFVSNDAPEDTGTLLELNALKDKDGFGRCKEGANPIDTP 1428
 |||||
 Db 1394 PPGGFPAITLHPMYGHQYTFSSMDERNNEHLEVLDVLYLLNRPGFGNCKLEEMLPDYP 1453

```

QY 1429 CQAGEEHTTAPYPTIMLDFONGNWTMONPSPACCCSDKIKKMLPVCPPACGLPPQ 1488
DB 1454 C-INATSMKTPSPSPNTHLPFOQKWTAAHPSPSCSCSREKLTMLPECPGAGARPQ 1512
QY 1489 RKONTADIILODLGRNIDSYLVKYVQIATKSLKNKIMVNERGSGFSGVNTQALPSP 1548
DB 1513 RIORSEVIELODLNRNIDSYLVKYVQIATKSLKNKIMVNERGSGFSGVNTQALPSP 1571
QY 1549 QEVNDAIKQKKHKLAKDSSADREINSIGR-----FMTGLDTRNN 1589
DB 1572 GE-----ALVGLSLGOMNNVSGGVTREASKEMLDLKHETTDN 1613
QY 1590 VKWENNGKMHAISSFLNYINNAILRANQKGENSHYITAPNPLNLTQOOLEVALM 1649
DB 1614 IKWENNGKMHAISSFLNYINNAILRANQKGENSHYITAPNPLNLTQOOLEVALM 1673
QY 1650 TTSVDVLSICVIFAMSEVPASFEVLIOERYSKAKHLOFISGVPEVLYMISNFWDMCN 1709
DB 1674 TTSVDVLSICVIFAMSEVPASFEVLIOERYSKAKHLOFISGVPEVLYMISNFWDMCN 1733
QY 1710 YVVPATLVIIHICQOQKSVSTNIPVLALLLXYGMSITPLMYPASFEVKIPSTAVV 1769
DB 1734 YVVPATLVIIHICQOQKSVSTNIPVLALLLXYGMSITPLMYPASFEVKIPSTAVV 1793
QY 1770 LTVSNLFIGINSVATFEVLEFTEFNK-LNNINDILKSVLEPHHCIGGLIDMKQNM 1828
DB 1794 LTVSNLFIGINSVATFEVLEFTEFNK-LNNINDILKSVLEPHHCIGGLIDMKQNM 1853
QY 1829 ADALEREGRNRPVPLSMVLGRNLEFAMAVEGVPEFLITVLIOYRFRIRPRVNAKLSPL 1888
DB 1854 ADALEREGRNRPVPLSMVLGRNLEFAMAVEGVPEFLITVLIOYRFRIRPRVNAKLSPL 1913
QY 1889 NDEDEVRERORIIDGGQNDILEIKELTKTYRKKRPAVDRIQVGIPEGCGELGVN 1948
DB 1914 NDEDEVRERORIIDGGQNDILEIKELTKTYRKKRPAVDRIQVGIPEGCGELGVN 1973
QY 1949 GAGKSTFKMLTGTVTYRGDAFLNKNLSINIHVNQMGCPQFQDAITELTGREHVE 2008
DB 1974 GAGKSTFKMLTGTVTYRGDAFLNKNLSINIHVNQMGCPQFQDAITELTGREHVE 2033
QY 2009 FPAALLRGVEKEVGVGEKATKGLVYKGEKAYAGNSGKNRKLSTAMALIGPVPVL 2068
DB 2034 FPAALLRGVEKEVGVGEKATKGLVYKGEKAYAGNSGKNRKLSTAMALIGPVPVL 2093
QY 2069 DEPTGMDEKARPLNCAALSVKEGRSVYLTSHSMECEALCTPMALMVNGRFGCLGSV 2128
DB 2094 DEPTGMDEKARPLNCAALSVKEGRSVYLTSHSMECEALCTPMALMVNGRFGCLGSV 2153
QY 2129 OHLKMPGDDGYIIVYRIAGSN---PDLKPYODFFGLAFPGSVLKEKHRNMLOQPSL 2184
DB 2154 OHLKMPGDDGYIIVYRIAGSN---PDLKPYODFFGLAFPGSVLKEKHRNMLOQPSL 2211
QY 2185 SSLATFSLISOSKKRLHEDYSQVSTLLDOYFVNFADQSDDHKLKQLSLH 2236
DB 2212 SSLATFSLISOSKKRLHEDYSQVSTLLDOYFVNFADQSDDHKLKQLSLH 2260

```

RESULT 3

002698

ID 002698 PRELIMINARY: PRT: 2281 AA.

AC 002698: 01-JUL-1997 (TREMBlrel. 04, Created)

DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE ABC transporter.

OS Bos taurus (Bovine).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

OC Bovidae; Bovinae; Bos.

OX NCBI_TaxID=9913;

RN [1]

RP SEQUENCE FROM N.A.

```

RC TISSUE=RETINAL ROD CELL;
RX MEDLINE=97248596; PubMed=9092582;
RA Iilling M., Molloy L.L., Molloy R.S.;
RT "The 220-kDa rim protein of retinal rod outer segments is a member of
the ABC transporter superfamily";
RL J. Biol. Chem. 272:10303-10310(1997).
DR EMBL: U90126; AAC48716.1; -.
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR003439; ABC_transport.
DR Pfam: PF00005; ABC_tran; 2.
DR Prodom: PD000006; ABC_transport; 2.
DR SMART: SM00382; AAA; 1.
DR TIGR: TIGR01257; rim_protein; 1.
KW ATP-binding.
SO SEQUENCE 2281 AA; 257228 MW; 71CD404C98F7A079 CRC64;

```

Query Match 49.7%; Score 5864.5; DB 6; Length 2281;

Best Local Similarity 50.1%; Pred. No. 0;

Matches 1158; Conservative 373; Mismatches 642; Indels 139; Gaps 26;

```

QY 6 QRLLLKMLTFRRRQNCOLLLEVAMPFTLLISVRLSYRPEEQHECHFPKAMPAG 65
DB 6 QRLLLKMLTFRRRQNCOLLLEVAMPFTLLISVRLSYRPEEQHECHFPKAMPAG 65
QY 66 TLPPVQGIICNANPCGRPTTPGEARGVGNFKSIYARLFSDARKLLYSOKOTSMKDM 125
DB 66 TLPPVQGIICNANPCGRPTTPGEARGVGNFKSIYARLFSDARKLLYSOKOTSMKDM 125
QY 66 MLPWLGIFPCNVNPPCFOSPAGEPQIVSNYNNSILARYRDFQELMDAPESQHLQGV 125
DB 66 MLPWLGIFPCNVNPPCFOSPAGEPQIVSNYNNSILARYRDFQELMDAPESQHLQGV 125
QY 126 RKVLRTLQOI-----KSSSNLKLQDFLVNDFEFGFGLYHNLSLPKSTYDKMLRAD 176
DB 126 RKVLRTLQOI-----KSSSNLKLQDFLVNDFEFGFGLYHNLSLPKSTYDKMLRAD 176
QY 126 WRELRTISQMLNTRLMRPERIAGRIIRIRVAKDDEMLTFLVKNIGLSNYSVYLLVNSQ 185
DB 126 WRELRTISQMLNTRLMRPERIAGRIIRIRVAKDDEMLTFLVKNIGLSNYSVYLLVNSQ 185
QY 177 YILKAVFLQGY-QLHLSL-CNGSKSEMIQLGDOEVS-----LGLPRKELAAERVL 229
DB 177 YILKAVFLQGY-QLHLSL-CNGSKSEMIQLGDOEVS-----LGLPRKELAAERVL 229
QY 186 V-REDFARGVPLMLKDIACSEALIERFLIPQRAAQVRSGLSLSQGLQWMDTL 244
DB 186 V-REDFARGVPLMLKDIACSEALIERFLIPQRAAQVRSGLSLSQGLQWMDTL 244
QY 230 RSNMDILKPLRTLNSPPEPSKEIAETKTLHLSIGTLAQLFMSRW-----SDMR--- 282
DB 230 RSNMDILKPLRTLNSPPEPSKEIAETKTLHLSIGTLAQLFMSRW-----SDMR--- 282
QY 245 YANVDK-----LHVPRRLDS-----RSGQMLNRSGRLISDMSPRI 284
DB 245 YANVDK-----LHVPRRLDS-----RSGQMLNRSGRLISDMSPRI 284
QY 283 -----QEVNLT---NVNSSSTQIYQVSRIVCGEGGLIKSLNMYEDNN 329
DB 283 -----QEVNLT---NVNSSSTQIYQVSRIVCGEGGLIKSLNMYEDNN 329
QY 285 QEFTHRSVQDILLVTRPLVQVOTGPEFQGLMGLDLCGYEGGSGVSPFNWYEDNN 344
DB 285 QEFTHRSVQDILLVTRPLVQVOTGPEFQGLMGLDLCGYEGGSGVSPFNWYEDNN 344
QY 330 YKALFGNGTEDEAFETFDNSTTPYCNOLMKNLSESPSLRIIMALKLPLLVGKIYTPPT 389
DB 330 YKALFGNGTEDEAFETFDNSTTPYCNOLMKNLSESPSLRIIMALKLPLLVGKIYTPPT 389
QY 345 YKAFVGLIDSTRKDIYSYDERTTFCNALQSLSPNPLRIKIMRAKAPLIMKGIITFPDS 404
DB 345 YKAFVGLIDSTRKDIYSYDERTTFCNALQSLSPNPLRIKIMRAKAPLIMKGIITFPDS 404
QY 390 PATROVAEYNAKTFQELAVFHDLEGMMEELSPKTIWFEMENSOEMDLVRLMDSRDNDHW 449
DB 390 PATROVAEYNAKTFQELAVFHDLEGMMEELSPKTIWFEMENSOEMDLVRLMDSRDNDHW 449
QY 405 PATRILIKMANSFEELERVRKLVKWEVEGPOIWFYFDKSTQMSMIRTLTENPPYKAFW 464
DB 405 PATRILIKMANSFEELERVRKLVKWEVEGPOIWFYFDKSTQMSMIRTLTENPPYKAFW 464
QY 450 EQLDGLDMTADQIVAFILAKHPEDVOSSNGSVYTWBEAFNETNOAIRTISRMECVNLK 509
DB 450 EQLDGLDMTADQIVAFILAKHPEDVOSSNGSVYTWBEAFNETNOAIRTISRMECVNLK 509
QY 465 NRQDGEGLTAEAVLNFVLYNGPREQADVDNENMDIFNIDRALRLANQYLECILDK 524
DB 465 NRQDGEGLTAEAVLNFVLYNGPREQADVDNENMDIFNIDRALRLANQYLECILDK 524
QY 510 LEPITAEVWLINKSMELDERKFMAGIVFTGTPGSLERHHVKKYIRMDINDVETNKI 569
DB 510 LEPITAEVWLINKSMELDERKFMAGIVFTGTPGSLERHHVKKYIRMDINDVETNKI 569
QY 525 FESYDDEFQULTORALSULEERNRFAVPRDMIPWTSPLRPHVKIKIRMDINDVETNKI 584
DB 525 FESYDDEFQULTORALSULEERNRFAVPRDMIPWTSPLRPHVKIKIRMDINDVETNKI 584
QY 570 KDGYWDPGRPADFEEDRMRYVWGFAVLQDVQDAITRVLTGTEKKTYGVMOQMPRCYVD 629
DB 570 KDGYWDPGRPADFEEDRMRYVWGFAVLQDVQDAITRVLTGTEKKTYGVMOQMPRCYVD 629
QY 585 KDRYWSGPRADVEDFEDFRYIMGEFAYLDQWVEHGRISQAOGEVPRGIGYLDQMPRCYVD 644
DB 585 KDRYWSGPRADVEDFEDFRYIMGEFAYLDQWVEHGRISQAOGEVPRGIGYLDQMPRCYVD 644
QY 630 DIFLRVMSRMPLEPLAMTYSAVYIKGIVYEKARKLETKRMINGLONSLIMESFSLSS 689
DB 630 DIFLRVMSRMPLEPLAMTYSAVYIKGIVYEKARKLETKRMINGLONSLIMESFSLSS 689
QY 645 DSPMILNRCFPLEFVNLAMYSVSMVTKSVILEKELRKETLKNQGVSNRYIWCWTFDLS 704
DB 645 DSPMILNRCFPLEFVNLAMYSVSMVTKSVILEKELRKETLKNQGVSNRYIWCWTFDLS 704
QY 690 LIPULVAGLIVILKLNGLNLPYSDPSVVEFVLSVAVYTILOCFILISFBRALMAAC 749
DB 690 LIPULVAGLIVILKLNGLNLPYSDPSVVEFVLSVAVYTILOCFILISFBRALMAAC 749
QY 705 FSLMSMSTICLITFIIMHGRIKLSNPFILFPLDLAFSIATIMQCFLLSFRASLAAAC 764
DB 705 FSLMSMSTICLITFIIMHGRIKLSNPFILFPLDLAFSIATIMQCFLLSFRASLAAAC 764
QY 750 GGIITFTLVPYVLCVAMQDYGFTLKIFASLSPVAFGCGCEYFALPEEQIGVQWMDL 809
DB 750 GGIITFTLVPYVLCVAMQDYGFTLKIFASLSPVAFGCGCEYFALPEEQIGVQWMDL 809

```

Db	765	SGVITFLYLPHILCFAMQDRIADKMMAVSLSPAFCGTETYLAFEDQGLQDSMN	824
Qy	810	FESPDEEGFNLTJISIMYLFDFLVGMVYIEAVEPGOYGIIPRWYPCJTSYWF--	867
Db	825	GNSPREDDEEFMSMKMLLDALYGLLAWYDOVPGOYGRPLWYFLQSSYWGCE	884
Qy	868	-----EESDEKSHPGSNQ-----KRMEFICMEEPETHKLQVSLQNLVXYRDSMKV	914
Db	885	GCSREERALEKTEPITEEMEDPEYREGINDCFERELPGJLVPGVCYKNVLIPEYGR	944
Qy	915	AVDGLALFEGQITSLFGLHGAGKTTMSLITGLPRTSGATYLKQJRKSEMTIRON	974
Db	945	AVDRNITFEYESQITAFGLGHGAGTTTSLTMTGLLPTSTSTVLVGKDIETNLDAIROS	1004
Qy	975	LGVCHONHVLFCMLTVEHIMEFAYALKLSEKHVAEMEOALDVGJLPSKLSKTSOLS	1034
Db	1005	LQMCQOHNLPHLHLTVAHILFLYQJLQKRSMDQAOLEMEALIEDGL--HNRKEARDLS	1063
Qy	1035	GMQOKLSVALAFVGSQKVLIDERTACVDYRSRGIMELLKTRQGTILLTSHHDEA	1094
Db	1064	GGVQRKLSVALAFVADAKVYUDEPTSGDYSRKSIMDLLKTRSGTITLSTHNDDEA	1123
Qy	1095	DVLGRIALITSHGKLQCVSSSLFLKNOJGTGYLLVWKDVSSLSQSNSSSTVSYLTK	1154
Db	1124	DILGDIALITISQRIYCSSTPYFLKNCFTGYLLVLR--KMTIOSQGRKATCSASK	1182
Qy	1155	EDSVQSSSDAGISGHESDTLTIVSAISLIRKHVSFARLVEDIGHETVYLPEAK	1214
Db	1183	GESVRCRPAQAEATTPROYLDG---DYNELTDWVHHNHRPAKLVEGIGELFLPLNKNFK	1239
Qy	1215	EGAFVELFHEIDRLSDLCISYGISSETTLEELFLKYAEESVDATPSGTLPARNRRA	1274
Db	1240	QRAYASIFRELEETLADLISGSFISDTPLEELFLKTVETDLSGHLFAGCTOQKEN---	1296
Qy	1275	FGQKOSTLR-----FTEDDADPNDSIDIDRESFTDLSGMDGKSYQVQGMKLTOQO	1328
Db	1297	INLHPCSGSEKAGOTPOQSSSHPEBPANHPRGQPPRRERGHISRLNS-----GARLTVGH	1352
Qy	1329	FVALLMKRLLIARRSKRGFFAQVILPAVEVICALVLSLIVPEFGKYSLELOPMYNEOY	1388
Db	1353	VOALLVRFQHTIRSHKDFLAQVILPATFVFLALMSLITPRGEXPALTLHPRMAYGOOY	1412
Qy	1389	TFVSNDAPEDTGTELELINALTKDREGSTRQMEGNNPIRDTPCQAGEEWTPAPYQIIML	1448
Db	1413	TFFSMDQPDSEBMLSALADVLVNRKFGFNRCILKEBMLPEPFC--GNSSPMKTPSVSDVTHL	1471
Qy	1449	FQNGNWTMONPSPACOSSDOKIKKMLVPCPGAGCLPPRQKONTADILDDTLGRNIDSY	1508
Db	1472	LQOQKWADQDPSPCRCSTREKTLMLTDECEBGAGGLPPRQIORSFTEIIDLDRVSDP	1531
Qy	1509	LKTVYVOITIAKSLNKIVMNEFRYGCFSLVSNQALPSPQEVNDAIKQMKHLKLAKDS	1568
Db	1532	LKVTYPALIRBSLSKFPVNNQRRGGISVG-----GKLPAFPPEGELV-----	1575
Qy	1569	SADRFNLISGR-----FMIGLDRNNVKKYFNNKGMHAIISFLNVI	1609
Db	1576	---GLSLDQLMNVSGGPMTRERAKEMPAFLKOLETEDIKNYFNNKGMHALVSFLNVA	1632
Qy	1610	NNAILRAVLQGENSHVGTAFANHPNLTLTQOOLSEVALMTSVDVYSICVIFAMSEVP	1669
Db	1633	HNAILIRASLHKRKNEEGTIVYISOPILNTEQJSEITVLTVSDAYVALCVIFAMSEVP	1692
Qy	1670	ASEVVELLOEVRYSKAKHLOFSGVSPVLYUMLSNEFMDMCNVTVPBATTIILFICQOKSY	1729
Db	1693	ASFVLYLIDERNKAKKHLOPVSQVSPITYWILNLFNEMIMNTVSAAIVGIFIGOKKAY	1752
Qy	1730	VSSNMLPVALLLILLYGMSITPLMTPASVEKEIPSTAVVLTSMFLFGINGSVATPYLE	1789
Db	1753	TJSEENPALVALLMLYGMVAPIMMYPASFLEPDIPTAVVALSCANLFGINGSSATPYLE	1812
Qy	1790	LFTDNK--LNNINDILKSVFLFPPHCCLCGLIDWYKNOAMADALEPFEENFVPSLMDL	1848

Db	1813	LEFNNTLRLINAMLRKLLIIPHCLGGLDIALSQAVDYNVQFGEAHSSNPQWDL	1872
Oy	1849	VGRNLFMAVEGVFFLITVLIOYREFIIRPRVNAKLSPLANDEDEDVREKQRIILDGGQ	1908
Db	1873	IGNLMAAAGVGVYFLITLLIOYOFFFSKWTTEPAKEPIITDEDDVAEEORIIISGNK	1932
Oy	1909	NDLLEIKELTYIRRRKRPVNDICVGIIPGDEFGILGNGAGKSSTFRKMLIGDTVTYG	1968
Db	1933	TDILRLNELTKVSSGTSPPVDRICVGRGDEFGILGNGAGKTTTKMLITGDPAVSG	1992
Oy	1969	DAFLKNSLTLSINEHYQNNQGYCPQFALITELTGREHEFFALLRGVPEKVGEGEMA	2028
Db	1993	DATVAGKSILINISIDVQMSGYCPQFALIDDLTGREHLYVARLGVPAEEIEBVTWMS	2052
Oy	2029	IRKGLIVYGEKVGANYSGCKNRKRLSTAMALICGPPVYLDEPTTMDKARFLMNCAL	2088
Db	2053	IQSLTSLIYADRLAGTYSGCKNRKRLSTALIALICGPLYLDEPTTMDQARRMLNTIM	2112
Oy	2089	SVYKGRSVLTLTSMCECEALCTRMAIMVNGFRCLGSVQHLKNFGGJYIVVRIAGS	2148
Db	2113	GLIREERAVLTLTSMCECEALCTRLIMVKGAFQCLGTIQLHKSFRFGGJYIVTKIRSP	2172
Oy	2149	N---PDLKPVODFEGALFAPSGVLTKEKRRNMLQYOLPSSLSLTARISFISIOSKKRLIE	2200
Db	2173	KDDLPLDGLPVDQFPGNPGPSQVGRERRHYNTLQGVSS--SLARIFRLVSHKSDLIE	2236
Oy	2205	DYSVSQTLIDQYFVNFAPKQSDDDHLKDLISH	2236
Db	2231	EVSVTOTLIDQYFVNFAPKQSD--TYDPLPLH	2259
RESULT 4			
ID	Q9BZC4	PRELIMINARY:	PRR: 2146 AA.
AC	Q9BZC4		
DT	01-JUN-2001 (TREMBLrel. 17, Created)		
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	ABC transporter member 7.		
GN	ABCA7.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RF	SEQUENCE FROM N.A.		
RX	MEDLINE=21328888; PubMed=11435699;		
RA	Brocardo C., Osorio J., Luciani M.-F., Schirml L.M., Prades C.,		
RA	Shulein S., Anouid I., Naudin L., Lafargue C., Rostler M., Jordan B.		
RA	Matei M.G., Dean M., Deneffe P., Chhim J.G.		
RT	"Comparative analysis of the promoter structure and genomic		
RT	organization of the human and mouse ABCA7 gene encoding a novel ABCA		
RT	transporter.";		
RL	Cytogenet. Cell Genet. 92:264-270(2001).		
RL	EMBL: AF328787; GenBank: U00959.1; "		
DR	InterPro: IPR003593; AAA_Atpase.		
DR	InterPro: IPR003439; ABC_transporter.		
DR	InterPro: IPR001899; Gram_pos_anchor.		
DR	Pfam: PF00005; ABC_tran; 2.		
DR	ProDom: PD000006; ABC_transport; 2.		
DR	SMART: SM00382; AAA; 2.		
DR	PROSITE: PS00211; ABC_TRANSPORTER; UNKNOWN_1.		
DR	PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.		
DR	ATP-binding.		
SEQUENCE	2146 AA: 234306 MW: 2391728055AD97E75 CRC64:		

Query Match	49.0%	Score 5775.5	DB 4	Length 2146
Best Local Similarity	50.3%	Pred. No. 0		
Matches 1148	Conservative 368	Mismatches 605	Indels 161	Gaps 26
QY	1	MACPOLRLILKNTLRRRCQCLLLVAMPLFLLILSVRLSYRPEYDNECHFNPKA	60	
db	1	MAFPTQMLILKNTLRRRCQCLLLVAMPLFLLILSVRLSYRPEYDNECHFNPKR	60	


```

Db 1369 POAVTSCGEVONTLGRNLSDFLVKTYPRVLVROGLTKKWNENYGGFSLG-GMDRLP 1427
OY 1547 PQGEVNDIAKOKKKHLAKADSSARFLNSIGREPTGIDTRNNKVPWNKGMHAISSFL 1606
Db 1428 SCQELGSRVEELMALLSPGLGADRLVRLKNTLMAHNSIDADDSLKIMWNKNGHSMVAFV 1487
OY 1607 NVINNAIILRANLQKGENPSHYGITAFNHPNLNTKOOLSEVNLMTSVYLVISTVIFAMS 1666
Db 1488 NKASAMILLRAHLPGRRAHNSITTLNPLNLTKEQLFEALMLKSSVDVLSTICVFRAMS 1547
OY 1667 FVPASFVPLIOERVSKAHLQFTISGVKPVITWLSNFWDMCNVVPATYIIIFIGEOQ 1726
Db 1548 FVPASFVPLVLEERVTAKHLQLMGSLPTLYLGNFLMDCNLYLPACIVLLEFLAQO 1607
OY 1727 KSYVSTNLPVALLLLYGSSTIPMPAPASFVKIPSTAVVLTSVLFGINGSVATF 1786
Db 1608 RAYVAPALPALLLLYGSSTIPMPAPASFVSPSTAVVLTICNLFIGINGSMATF 1667
OY 1787 VLEFTDNKLNINDILKSVPLIFPHFCLGRLIDMVKNQAMADALERGFNFPVPLSW 1846
Db 1668 VLEFSDOKLOEVSRLKQVFLIEPHFCLGRLIDMVKNQAMADALERGFROFOSPLRW 1727
OY 1847 DLVGNLFAMAVEGVFPLITVLIQYRFFTRPRVNAKLSPLNDEDEDVRRERQILDG 1906
Db 1728 EVVGKNTLAMYIQGFLFTLLQHSQILDPQPRVNSTPLGEBEDDVAREFRVVOGA 1787
OY 1907 GONDILKELTKIYRRKRKPAVDRIQVIRPGECFGLIGNAGKSTKMLTGDTTV 1966
Db 1788 TQGVPLVRLNTIKYRGCRMVAVRDLGLGPGECFGLIGNAGKSTKRMVYGDVLAS 1847
OY 1967 RGDAFLKNSLISNIEHYQNMGYCPQFDATITELLTGREHEFPALRGVPEKGVKGE 2026
Db 1848 RGEAVLAGHSVAREPSSAHLMSGYCQSDALFEELLTGREHELLRLRGVPEAQVOTAG 1907
OY 2027 WAKRLGLVYKGEYAGVSGNKRKLTSTAMALLGPPVFLDEPTTGMPKARFPLMNC 2086
Db 1908 SGLARLGISWYADRPAGYSGGNKRKLTATLALVGDPAVVFLDEPTTGMPDSARFLMNS 1967
OY 2087 ALSVYKGRSVVLTSHSMECEALCTRMALMVNGRFGCLSVQHLKRFEDGTIVYRIA 2146
Db 1968 LLAVVREGRSVMTLSHMECEALCSRLAMVNGRFGCLSPQHLKRFAGHTLTIRVP 2027
OY 2147 GSNPDLKPVQDFGLAPPGSVLAKKHKHNMLOYL-PSLSLARIIESILSOSKRLIHED 2205
Db 2028 AARS--OPAAAFVAAEPGSGELREAHGRLRFQLPFGRCALARVGEGLAVHGAHEGVED 2085
OY 2206 YVSQTLTLDQVFNFAKDQSDDDLKDLSLHKQTY-VDVAV-----LTSLODEKXK 2257
Db 2086 FSVSQTLMEVFLYFSKQDQKDE---DTEQCKRAGVGVDAFGLQHPKRVSOFLDPESTA 2142
OY 2258 ES 2259
Db 2143 ET 2144

```

```

RA Broccardo C., Osorio J., Luciano M.-F., Schriml L.M., Prades C.,
RA Shulman S., Arnold I., Naudin L., Lafaque C., Rostier M., Jordan B.,
RA Mattei M.G., Dean M., Denelle P., Chimalin G.,
RT "Comparative analysis of the promoter structure and genomic
RT organization of the human and mouse ABCA7 gene encoding a novel ABCA
RT transporter."
RL Cytogenet. Cell Genet. 92:264-270(2001).
DR EMBL; AF287142; AAK56863.1; -.
DR EMBL; AF287141; AAK56862.1; -.
DR MGI; MGI:1351646; Abca7.
DR InterPro; IPR003439; ABC_transport.
DR InterPro; IPR002016; Peroxidase.
DR Pfam; PF00005; ABC_tran.2.
DR ProDom; PD000006; ABC_transporter; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; UNKNOWN_1.
DR PROSITE; PS00435; PEROXIDASE_1; UNKNOWN_1.
KW ATP-binding.
SQ SEQUENCE 2159 AA; 236882 MW; CD2BE3FE0B8B822B CRC64;

Query Match 48.6%; Score 5732; DB 11; Length 2159;
Best Local Similarity 50.6%; Pred. No. 0;
Matches 1144; Conservative 362; Mismatches 581; Indels 176; Gaps 31;

OY 1 MACWQLRLMLKNLFRFRROTCOLLLEVAMPFLITLISVRLSYRPEDEHECHPKKA 60
Db 1 MALGTOLMLLKNNTYRRROPIQLLVLLWPLFLIIVAAHSHPLHEHCHPKKP 60
OY 61 MSAGTLPVVQGITCANNPCEFRPTPGEPAGVGVNGFNKSIYARLFSDARLLLYSQKDT 120
Db 61 LPSAGTVPWLOGVGNVNSCFQHPPTGPKPGVLSNFQDSLIRLLADTRVYL----GGH 116
OY 121 SKMDMKVYRTQQLKSSSNLKLQDELVDNTEFSGFLYHNLISLPKSYDKMLRADVYLH 180
Db 117 STQMDLALGLKILPV----- 131
OY 181 KYFLQGYQLHTSLQNGKSEEMIQDQEVSELCGLPKREKLAARVY-RSNMDILKPI 239
Db 132 -----LRAVGAGCARPQE-----SDQPTSQ--GSVTKLL--EKLQKASLD---PV 169
OY 240 LKTLNSTSPFSPKLEALAKITLLHSLGLAQLPLSMRSWSDMRQEVMLFVNSSSSSTQ 299
Db 170 L-----GQADSSMRKFSDAIRDLAQLILPLPSIMELR--ALLRPPRSAGSL 215
OY 300 IYQAVSRIVC--GHPREGGLKTRKSLNMYEDNNYKALFSGNGTEDEATFTYDNSTTPYCN 356
Db 216 L---VSEALCSTKQSSSGGL---SLNMYEANOINERPG---PEVAPALPDNSLSPACS 265
OY 357 DLMKLESSPLSRIITWAKLPLVLGKILYTPDTPATQVMAEVNKTQELAVFHDLEGMW 416
Db 266 EFVGTLDLDDHPVSRLLMRRLKPLILGKILFAPDTNPTKRLMAQVNTQFBEALALLDLHLW 325
OY 417 EELSPKWTFMENSGEMDLVRLMLDSRNDHFWRQDLDGMLTADQIYAFIAKIPEDVQS 476
Db 326 GYLGPQITFENFMDSTNVMLQLDLVGGTGO--RQGPRAOKKLEALIKDFL-----DP 376
OY 477 SNGSVYTWREAFNETNOAIRTISRFMECVNLKLEPIATEYWLINKSMEFLDERKEFWAGI 536
Db 377 SRGG-YSMREAHADMGRLAGILGQMECVSLDKLEAVSEBALVSRALELDEGRRLAGI 435
OY 537 VFTG-----ITPGSTF---LPHVAKYKTRMDIDNVERTNKTIKDGYWDPGRPADFEEDMY 588
Db 436 VFLSPHPLDSELSPPALSPGHRLFRKIRMDIDVTTRNKRIRDFKMPDGPSPADPFMDLRY 495
OY 589 VMGGAFLQDQVVEOALILVLTGTEKKGTVYVQOMPYCYUDDTFLRWSSMPLFMTLAW 648
Db 496 VMGGAFLVQLDLEQAAVAVLGGNSRGLYIQQPHHCYVDDVFLVRLVLSLPLETLAW 555
OY 649 IYSAVAILKGIYVEKEAKLKTMRITMGDLSILWFWSFISLPLVLSAGILAVYIKGN 708
Db 556 IYSAVLVYKAVREKETRLRETMRAMGISRAVVLWLGWFLSGIGFLVSALVLYLKLGN 615
OY 709 LLFYSDBSVAVVSVRAVVTITLQCFILSTLFSRANLAAACGGITTYTLYLVYLVCAWQ 768

```



```

Db 616 ILPSHPVIELELAAFAVAATVAOSFLLSAFSPSRANLAAACGLAYFALYLPYLVCVAMR 675
Qy 769 DYVGFELKIFASLLSPVAFGCECFALFEEOGIGVOMDNLPESEVEDGPNLTJTSIMM 828
Db 676 ERLHLGGLAALLSPVAFGCESEALLEEODGDAOHNLGTGA-EDVSLAOVSAFL 734
Qy 829 LFDTELYGVMTYEAVERPGYGIIPRMYPCTKSYWFGESDEKSHGSONKRMSELCM 888
Db 735 LIDAVYGLATWLEAVECPGQYGIPEPMNFPFRSYWCG-PGPRSSVLAAPQDPKYLIV 793
Qy 889 EEPERHLKIGVSIQNLVKKYRDMKAVADGLALNFEGQITSFELCHNGAGKTTMSITG 948
Db 794 EEPRLGLVGVASIRGLKHKFRGCCPQALOGMLNDEYEGHITPAIHOHNGAKTTLSITG 853
Qy 949 LPPPTGTAYILGKDIRSEMTITRONLGVCPQHNVLFDMLTVEEHIMYARLKGSEKHV 1008
Db 854 LPPSSGASIIIGHVQTMALIRHLGICPOYNVLFMLTYEEHVMYGRKKGSAAM 913
Qy 1009 KAEOMALDVLPSKSKTSKTSQSGMQRKLSVALAFVSGSKVILDEPTAGVPYSR 1068
Db 914 GPERELIRDVGL-TLKRDTOTRHLSGMQRKLSVALAFVSGSRVYIINDEPTAGVPSR 972
Qy 1069 KRIWELLKYRGRTIITSTHMDADVIGRIALISHKGLCCVSSLPKNOIGTYL 1128
Db 973 KRIWELLKYRGRTIITSTHMDADVIGRIALISHKGLCCVSSLPKNOIGTYL 1032
Qy 1129 TLVKKDVESLSSCRSSSTVSYLKKEDSVSSQSSDAGISGDSHSDT----- 1175
Db 1033 TLVK-----SSQSLYTHDAKGDSDEPRREKSKSGMQRKTSIDTAFTRGTSKSNQA 1081
Qy 1176 -----LTIDVSAISNLRKHVSEARLVEDIGHLELYVLYPEAKEGAFVELFHEIDR 1228
Db 1082 PAPGAVPIPTSTARILELVOQHVPAGQLYVEDLPHELLLYLPYAGALDSEFAMVPELDOQ 1141
Qy 1229 LSDIGSSYGETLEETLEELFKVAESGVADATSDGTPARANKRAFGDKOSCL-RPPTG 1287
Db 1142 LELGLTGIGIDTLEELFKLVNDAHEGDSRPDLHLR-----TCTPQPTG 1191
Qy 1288 DDAADPNDSIDIPESRENDLISGMKGSKYQVKGWKLTOQOQFALLMKRLILARRSRGF 1347
Db 1192 PRASVLENGELAPQ-----GL-APNAQYQGWTLTCQOQFALLMKRLILARRSRGL 1242
Qy 1348 PAQIYLPAYFVIALVFSILVPECKYSLLEQPMWYEDQYTFVSNDAPEDTGILELNA 1407
Db 1243 PAQVYLPAYFVIALVFSILVPECKYSLLEQPMWYEDQYTFVSNDAPEDTGILELNA 1302
Qy 1408 LKKDGFGRKCEGNPIPTPCQAGEEEMTTPR-VPTIIMDFQNGMTMKNPSACCS 1466
Db 1303 LKGEAGLOEPSQMDKADARSECTHSLACFTYVPEYPPVASTILASGNMTPESPACCS 1362
Qy 1467 SDKIKMLPVCPPAGGLPPQORKONTADIIODLGRNISDYLVTYQIITAKSLKRTW 1526
Db 1363 QGARRLLPDGACGAGPPQAVAGLEGVONLGRNVSDLVKTYISVARGLKTKW 1422
Qy 1527 VAEFRYGGSLGVNTQALPESQEVNDAIKMKKILAKADSSADRTINSLGREMTGLDT 1586
Db 1423 VAEFRYGGSLGVNTQALPESQEVNDAIKMKKILAKADSSADRTINSLGREMTGLDT 1481
Qy 1587 RNNVYVFNKMGMAISSFLYINNAIIRANQKGENSHGITAIFNPIMLTQOJSEV 1646
Db 1482 RNSLKIWNKMHMAVAVANRANGLHALPSPGVRRHAISITTLNPLNLTKEQJSEA 1541
Qy 1647 ALMTTSVDLVISCIYFAMSPVAFVFLIOERYSKAKHLOFISGVFVYIWLNSFNWD 1706
Db 1542 TLASSVDLVISCIYFAMSPVAFVFLIOERYSKAKHLOFISGVFVYIWLNSFNWD 1601
Qy 1707 MCNYVVPATVITFCQOKSYSTNULPVALLLLYGSITPPLMYPASFEVEKIPSTA 1766
Db 1603 MCNYVVPATVITFCQOKSYSTNULPVALLLLYGSITPPLMYPASFEVEKIPSTA 1661
Qy 1767 YVVLTSVNLFIGINSVATFVELFTDNKLNININDILKSVFLPHFCLGGLIDMVNO 1826
Db 1662 YVVLTSVNLFIGINSVATFVELFTDNKLNININDILKSVFLPHFCLGGLIDMVNO 1721

```

```

Qy 1827 AMADLERFGENREVPSLMDLVGRNLFAMAVGVPEFLITVLIOYREFIRPRPVNAKUS 1886
Db 1722 AMAAFERLGRKQFOSPLRNDIICKNLAMAGCPLELITLLQHNRLLPQSKRLLP 1781
Qy 1887 PLNDEDEVRERQRILLDGGQNDILEIKELTKYRRKRPAYDRICVGPPECEGLG 1946
Db 1782 PLGEDEDAQERBRVATKGATQGVVLRLDTLVYRGQRPAVDRLCTGIPPECEGLG 1841
Qy 1947 VNGAGKSTFRMLTGDTTVTRGDAFLNKSILNSIHEHONMKYCPQFDAITELTGREH 2006
Db 1842 VNGAGKSTFRMLTGDTTVTRGDAFLNKSILNSIHEHONMKYCPQFDAITELTGREH 1901
Qy 2007 VEPFALLRGVEPEKVGVEWEMAIKRLGLVYGEYAGNYSGKKRKLSTMALJIGPPV 2066
Db 1902 LELARLRGVPEAOVATOLSLGLVRLGPSTADRPATYSGGKRLATALVGPDAV 1961
Qy 2067 FLDEPTTGMDPKARRPLMNCALSVKEGRSVLTSHMECECALCTRMATMVGRCPLG 2126
Db 1962 FLDEPTTGMDPKARRPLMNCALSVKEGRSVLTSHMECECALCTRMATMVGRCPLG 2021
Qy 2127 SVQHLKRNFGDGTITVYRIAGSNPDLKPVDFGLAFPGSVLKEKIRNMQYOL-PSLS 2185
Db 2022 SSQHLKRNFGDGTITVYRIAGSNPDLKPVDFGLAFPGSVLKEKIRNMQYOL-PSLS 2079
Qy 2186 SLARFSLISQSKRKLHIEDSVSQTLDOVFVNAFADQSDDD 2228
Db 2080 TLTVFRELAAQGRHAHVEDFSQTLLEEVFLTFSSKDGSEB 2122

RESULT 7
ID 096S58 PRELIMINARY; PRT; 2008 AA.
AC 096S58;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE ABCA-SSN.
GN ABCA7/ABCA-SSN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2155283; PubMed=11355874;
RA Tanaka A., Ikeda Y., Abe-Dohmae S., Atakawa R., Sadanami K.,
RA Yokoyama S., Ueda K.,
RA Kikawa A., Nakagawa S., Nagase T., Aoki R., Kioka N., Amachi T.,
RT "Human ABCA1 Contains a Large Amino-Terminal Extracellular Domain
RT Homologous to an Epitope of Sjogren's Syndrome";
RL Biochem. Biophys. Res. Commun. 283:1019-1025(2001).
DR EMBL: AB055390; BAB62294.1; -.
DR InterPro: IPR003439; ABC_tran;
DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam: PF00005; ABC_tran; 2.
DR ProDom: PD000006; ABC_transporter; 2.
DR PROSITE: PS00211; ABC_TRANSPORTER; UNKNOWN_1.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
SQ SEQUENCE 2008 AA: 218617 MW: 226F85C24230B90 CRC64;

Query Match 45.6%; Score 5376; DB 4; Length 2008;
Best Local Similarity 52.2%; Pred. No. 0;
Matches 1065; Conservative 334; Mismatches 554; Indels 88; Gaps 23;
Qy 247 SPFSKELA-----EATKTLHSIGTLAQLFMSRSMKROEYMLTVNNSSSSSTQIY 301
Db 26 SPFSKELAQLAQAQEPHLSLEAEADLAQELLRLSLVEYR---ALLQPRGSGPLEL 81
Qy 302 QAVSRIVCGHEEGGGLKIKSLNWEYEDNNYKALFGNGNTEDEATFTFDNSTTPPCNDLMKN 361
Db 82 --LSEALCSVRGSPSTVGPISLWYEAJDIMELVG---QEPESALPDSSLSIPACSELIGA 135

```


QY 362 LESSPLSLIMKALKEPLVGLILYTPDPATROVAEYNKTEFOELAVPHLEGMWELSP 421
 136 LSHPLSLRILMRKILGLIKLLFLPADPTFKKLAQVNRTEFEELTLDRVREVMELGP 195
 QY 422 KIMTFENSOEDVLVIMLDSRDNHFMFOQLDGLDWTMADIVAFLLAKHPELVOS----- 476
 196 RLTTEWNSNVAMLORLLOMODEGR-RORPGRD-----HMEALRSFLDPG 242
 QY 477 SNGSVTWREAFNETNOAIRITISRECEVNLKLEPIATEVWLIMKSMELDEREFMAGI 536
 243 SGG--YSMDANADVGHVLGTGRVTEGLSLDKLEAPSEALVSRALQOLLAHEHFMVAV 300
 QY 537 VPLGITPGSELPH-----HWKTRMIDONVBRINKIDGWDEGPAPDEPMRY 588
 301 VPLGPEDSSDPTEHPDPLGPHVRIKIMDIDVYTRNKIRDRWMDGPAPADPLTLRY 360
 QY 589 VMGGEFYLDVYEOALIRVLGTGEKTKGYVMQOMPYPGVDDIPLRVMSRSPMLMTLAW 648
 361 VMGGEFYLDVLEBAVRVLSGANPRAGLYDOOMEPGVDDVPLRVLSRSLPLTLTAW 420
 QY 649 IYSVAVINGIYKEARLEKEMTGMJDNLSIMFWSPLSLPLVLSAGLLVILKLG 708
 421 IYSVTLVAVVREKETRLRDTMRAMGSRVAVLIMGLFSLCLGPELLSAILVILKLG 480
 QY 709 LILYSDPSVVEVLSYAVVTLLOCFILISTESKRAMLAACGGIIFLYLYLVYCVAMQ 768
 481 ILPSHGVVFLFAFAVATVQSFLLSAFSSRANLAACGLAYFSLTYLVYCVAMR 540
 QY 769 DYVGFILKIFASLLSPVAFGECEYFALFEEOGIGVOMNLFESVVEEGEGLTISIM 828
 541 DRIPACGRVAAISLSPVAFGECEYFALFEEOGIGVOMNINVTGRT-ADVFSLAOGSL 599
 QY 829 LEDTLYGVWTVLEAVFPGYIGIPREVPFCTKSTWFEESDEKSHPSQNMKESEICM 888
 600 LDDAALYGLATVLEAVCGOYGIPEPMNFPFRSSWCPRPKSPACPPPLD-PKVLV 658
 QY 889 EEPRTLTKGVSTONVVKYRDGMKAVYDGLALNFYEGDITSELGHNAAGTKTMSILTG 948
 659 EEPRTLSGVSVRSLEKFPSSPOLNGLSLDPTOQHITAFLOHNGAKTTLSILSG 718
 QY 949 LEPTSGTAYILGKDIRSEMSTIRONLGYCPQHNLYFDMALVEEIMWYARLKSEKHV 1006
 719 LEPTSGSAFIIIGHDVRSSMAIRPHLGYCPQYNVLEDMALVDEHVMYFGRKKGSAVV 778
 QY 1009 KAEWEMALDVGLPSSKLSKTSQLSGGMOKLSVALAVGSSKVVILDEPTAGVDPISR 1068
 779 GPPQDRLQDVG-LVSKQSVQTRHLSGMOKRLSVAIAFAVGGSOYVILDEPTAGVDPISR 837
 QY 1069 RGIWELLKYRGRTIILSTHMDADVLDGRITAIISHKLCYSSSLFLKMQLGTYL 1128
 838 RGIWELLKYRGRTIILSTHMDADVLDGRITAIISHKLCYSSSLFLKMQLGTYL 897
 QY 1129 TLVKKDVESLSSCRNSSSTVYLKEDSVSSSSDAGLGDHESDTLIDVSAISNLIR 1188
 898 TLVKKALPLTJN-----EKADTDMGVSVDTRQEKKNQSGSRSVGPQLALVQ 945
 QY 1189 KHVSEARIVEDIHGLTYVLYEAKKEGAFVLFHEIDRLSDLSISGISSETLLEIF 1248
 946 HWPGRARLEVEELPHLVLVPLTGAHDGSFATLFRDLRLAELRLTGCSIDTLEIF 1005
 QY 1249 LKVAEESGVDAETSDGTLPPARRNRAGF-DKOSCLRPTEEDAAPNSDIDPESRETL 1307
 1006 LKVAEESGVDAETSDGTLPPARRNRAGF-DKOSCLRPTEEDAAPNSDIDPESRETL 1063
 QY 1308 LSGMDCKSGYQVGMKLTQOQFVALLMKRLIARRSRKGFQIYLVPAVFCIALVFSLI 1367
 1064 GSGPAPVAVG--RVQGMALVROQALALLKRELLARRSRKGFQIYLVPAVFCIALVFSLI 1121
 QY 1368 VPEFGKYSLELQPMWYNQYTFVSNDAEDDTGLELLNALTKDGFGRCHMGAPIPDT 1427
 1122 VPEFGKYSLELQPMWYNQYTFVSNDAEDDTGLELLNALTKDGFGRCHMGAPIPDT 1171
 QY 1428 PCQAGEEEMTAPVDTQINDLFQNGNMTWQNSPACQCSSDKIKRMLPVCPPGAGGLPPP 1487

Db 1172 PVQSHRSFAPAEVPAEVARVLAGSNMTPESSPACQSQPGARRLPLPCPAAGGPPP 1231
 QY 1488 QORQNTADILLODLGRNISYLVKTYVOITAKSLKXKINWERNRGGSGISVNTQALPP 1547
 Db 1232 QAVTSGSEVONLGRNLSPLVKTYPRLOQGLKTKKMWENRGGSGISG-GRDGLDS 1290
 QY 1548 SOEVNDAIKOMKHLKLAQSSADREFLNSLGRFMTGIDTFNNKXVFNKNGHAIISFLN 1607
 Db 1291 GQELGRVVEELMALLSPRLPGALDRVLKNTLVAMHSLDAQDSLKIMFNKNGHMSVAFVN 1350
 QY 1608 VTNATILRANIOKGENSHYGTAFNHPNLTLTQOQSEVALMTTSVDVLSICVIFAMSE 1667
 Db 1351 RASNAIILRAHLPGRARHSHITTLNHPNMLTKEOLEBAALMASSVDVLSICVIFAMSE 1410
 QY 1668 VPASFVFLIOERYSKAKHLQFISGVKPVLYLMSNFVMDCNVAVPALYIIIFICEFOOK 1727
 Db 1411 VPASFVFLIOERYSKAKHLQFISGVKPVLYLMSNFVMDCNVAVPALYIIIFICEFOOK 1470
 QY 1728 SYVSTNLPVILALLLLYGSITPLMYPASVFERIPSTAVVLTSVNLEIGNSVATEV 1787
 Db 1471 AYVAPANLPALLLLLLYGSITPLMYPASVFERIPSTAVVLTSVNLEIGNSVATEV 1530
 QY 1788 LEFTDKLNNINDILKSVFLIPPHCGRLIDMKYKQAMALDRPGENRVSPLSMD 1847
 Db 1531 LEFTDKLNNINDILKSVFLIPPHCGRLIDMKYKQAMALDRPGENRVSPLSMD 1590
 QY 1848 LVGRNLPMAVEGVFFLTLVLOYRFFIPRPVNAKLSPLNDEDEDRERORILDDGG 1907
 Db 1591 VVGNLNLAMVIGQPLFLTLVLOYRFFIPRPVNAKLSPLNDEDEDRERORILDDGG 1650
 QY 1908 QNDILEIKELTKYRKRRAVDRICVITPPECFGLLGVNGAKSSTFKMLTGDTTVR 1967
 Db 1651 QGDVLYRNLTAKYRGORMPAVDRICVITPPECFGLLGVNGAKSSTFKMLTGDTTVR 1710
 QY 1968 GDAPLKNLSILNHEVHOMGYCOPDATTLELTGHEHEFFALLRGVPEKVGKVGSM 2027
 Db 1711 GDAPLKNLSILNHEVHOMGYCOPDATTLELTGHEHEFFALLRGVPEKVGKVGSM 1770
 QY 2028 AIRRLGLVKYGEKAYAGNYSKGNRKRLSTAMALIGPPVFLDEPTGMDPKARFLMCA 2087
 Db 1771 GLARGLSWADRPAGTYSKGNRKRLSTAMALIGPPVFLDEPTGMDPKARFLMCA 1830
 QY 2088 LSVYKEGRSVYVLSHMEDEALCTRAMVNGRFRCLGVOHLKMRPGDGTIVYRIAG 2147
 Db 1831 LSVYKEGRSVYVLSHMEDEALCTRAMVNGRFRCLGVOHLKMRPGDGTIVYRIAG 1890
 QY 2148 SNPLKPVODFFGLAPFSGVYLKEHHRMLQOY-PSLSIARIFSLISQSKRLHIEDY 2206
 Db 1891 ARS--QPAAPFVAAPFSGVYLKEHHRMLQOY-PSLSIARIFSLISQSKRLHIEDY 1948
 QY 2207 SVSOTLLQYFVNFAPKQOSDDHLKDLSLHKNOTV-VDVAV-----LTSFLQDEKVE 2258
 Db 1949 SVSOTLLQYFVNFAPKQOSDDHLKDLSLHKNOTV-VDVAV-----LTSFLQDEKVE 2005
 QY 2259 S 2259
 Db 2006 T 2006

RESULT 8

Q9ESR9 PRELIMINARY; PRT: 2434 AA.

AC Q9ESR9; 01-MAR-2001 (Tremblrel, 16, Created)
 DT 01-MAR-2001 (Tremblrel, 16, last sequence update)
 DE ABC2. 01-JUN-2002 (Tremblrel, 21, last annotation update)
 GN ABC2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_taxid=10116;

RN [1]
 RP SEQUENCE FROM N.A.
 RX TISSUE=BRAIN;
 MEDLINE=20427113; PubMed=10970803;
 RA Zhao L., Zhou C., Tanaka A., Nakata M., Hirabayashi T., Amachi T.,
 Shioda S., Ueda K., Inagaki N.;
 RT "Cloning, characterization and tissue distribution of the rat ATP-
 binding cassette (ABC) transporter ABC2/ABCA2.";
 RL Biochem. J. 350:865-872(2000).
 DR EMBL: AB037937; BAB16596.1;
 DR InterPro: IPR003593; AAA ATPase.
 DR InterPro: IPR003439; ABC transporter.
 DR InterPro: IPR000561; EGF-like.
 DR Pfam: PF00005; ABC_tran; 2.
 DR Prodom: PD000006; ABC_transport; 2.
 DR SMART: SM00382; AAA; 2.
 DR PROSITE: PS00211; ABC_TRANSPORTER; UNKNOWN_1.
 DR PROSITE: PS00022; EGF_1; UNKNOWN_1.
 DR PROSITE: PS00213; LIPOCALIN; UNKNOWN_1.
 KW ATP-binding.
 SQ SEQUENCE 2434 AA; 270925 MW; CD424A9C4F63513F CRC64;
 Query Match 36.0%; Score 4249.5; DB 11; Length 2434;
 Best Local Similarity 39.1%; Pred. No. 5.1e-270; Indels 483; Gaps 60;
 Matches 992; Conservative 346; Mismatches 713;
 6 QRLRLMKMLTERRRQTCOLLLEVAMPLEFILILSVLSYPPYEQHCHFPKAMPAG 65
 6 QIOLLMKMTVLRKSRPWLAFELFIPVLFIILGLRQKPTISVKEAFYAAFLTSAG 65
 66 TLPWOGITCNANPCFRPTPEGEAPVGVGNKSIIVARLFDARLLYSOK -DTSKD 124
 66 ILPVMOSTLCPDQORDEFFGL-----QYANSTVTLERLKRKYEESMLFPERPS 115
 125 MKRYLTLO-----IKKSSNMKLQDFLVNENFSGLYNLSLPKST 168
 116 LQSELEALHQRLEALSSGGTWEHSARPAVSSFLDSVARCKRELFLMONLSPNST 175
 169 VDKMLRADY----- 177
 176 AQAALLAARVDSEFVYRLFGPLRDLDGKLGFLRKQEPMSHLSNPLFQWELLAPALLE 235
 178 -----IL-----HKVFLQGYQLHLTSLCNGSKSEMIQLGQEVSELGLPK 219
 236 QLTCAFGSELGRILTMPGHOVDLQYR---DAVCSGQATARA---QHESDL----- 282
 220 EKLAARVLRSMQDLKRLTL-----NSTSPRSKELAEATKLLHSLGLAOELFS 274
 283 ----ATE--LRNOLDIAK-IAQQLGNVPNGSPQFOAPSPQSLALGLDLDLVQKVLQD 335
 275 MRSMSDMROEVMFLTVNNSSSSTOYQAVSRIVCGHPREGGLIKTSLMWEDNNYKALF 334
 336 V---DVLGALLLPLQGCACAGRAPQA-----GSPGSPA-----NSTGVG 373
 335 GCGTGEDAETFDNSTTTPYCDIMKLNLESSPLSRITIKALPLVG----- 381
 374 ANTPVTVEEGTQSPVTPASPDITQGCASAFYO--LMAGLOPILCGNRTIEPEALRG 431
 382 -----KILYTPPTPATROVMAEVNKTFOELAFHLEGM 415
 432 MSSSLGFTSEQNLGLLHMLTSNKILYAPAGSEADHILKANETFAVGNVTHYQV 491
 416 WEELSPRIWTMENSOEMDLVRMLDSRDNDHFWEQOLDGLMTAADIYAFKAPEDVQ 475
 492 WMTISAEIRSFLE-----QGRLOQH-----LHW-LQGYVADLRLHPEANM 530
 476 SS-----NGSVYTWREAFNETQALRTIS-----RMECNLKLKLP 512
 531 LSLDELPPALRLDYFSLPNGTALL-----QOLDITIDNACGMIOGMSKVSVDIFKG 581
 513 IATEVWLINKSME--LLDERKFWAGIVFTGTPGSTIELPHHYKTKIRMDIDNVERINKIK 570

Db 582 FPDSEIANTVLTNOAYQADNVTFASVIFQTRKDS--LPPHHYKLRONSSEFTEKTEIR 639
 QY 571 DGWDCPRADPFEDMRKYWGFAVLODYVEQAIIRVLGTE-KRTGVYMQMPRYCYVD 629
 Db 640 RAYWRPBPNTG---GREYFLYGFVWIODMTERAIINTFGHVDVEGNAVYOMPRYCYTR 696
 QY 630 DPLRVSRMPLFMTLVAIVSAVAITKGIYKEERLKEKTRMIGDINSILMFSPFIS 689
 Db 697 DDFLEVEHMPICMTVSWY SVAMTQIHVAERHLEKRWMTKGLNNAHVNAVFITG 756
 QY 690 LIPPLVAGLIVYLKLGNIPLPYSDPSVFEPLSVAVVITLQCFLLSTPLSRANLAAC 749
 Db 757 FVQLSIVTALTLKALGYGLMHSVLLIWLFLAVYAVATIMCFLVSYLSKALIASC 816
 QY 750 GGIYFTLLPYVIC-----VANODYGFLLKIFASLISVAVAGFCGCFALFEEDGIV 804
 Db 817 GGIYFLSYVYVYVAVVIAIREVADHKITAFE-KCIASLMTSTAGLSKYFALEVAGVGI 875
 QY 805 QMDLFPSPVEEDGFNLITSLIMLFDYPLVGYWMTVIEAVPFGOYGPFPCTKSY 864
 Db 876 QMTFFSOSPVGEDDFNLLAVTLMVDTVYVGVITWYIEAVHGMGMLPRPWYFPIQKS 935
 QY 865 WFGD---ESDEKSHPGSNOKRMS-----ETC-----MEEPYHLKLGVSIOQL 904
 Db 936 WLGSGRTETWMSWPMNAHAPRLSVMEEDQACAMBSRHEETRGMEERTPLPLVGVCKL 995
 QY 905 VKVYRDGKVAVDGLANFTFGQITFSLGNGAGKTTYSILGLPPTSGTAYILCKDI 964
 Db 996 TKVYKKNKIALNLSTLNTYENOVSEFLGNGAGKTTWILGLPPTSGSATYIGHDI 1055
 QY 965 RSEMTIRONLGVCPQHNVLFDMLTVEEHIMFYARLKLSEKHAKYAKMEDALDVGIPSS 1024
 Db 1056 RTEMDIRKNLGKCPQHNVLFDQLTVEEHIMFYSLKSMQOBEIRKMDKMTEDLE-SN 1114
 QY 1025 KLSKTSQLSGMOQRKLSVALAFVGGSKVYLDEPTGAVDPYSRGIWELLKYRGRTI 1084
 Db 1115 KRHSIYQTLGSGMKRKLISVALAFVGGSRATILDEPTGAVDPYARRAIMDLILYKPKRTI 1174
 QY 1085 ILSTHMDADYLDGRIAIISHGKLCVGSSEFLKNDLGTGYITLYKKVDSESLSCRN 1144
 Db 1175 LLSTHMDADLDDRIALIIISHGKLCGSPFLKGAYGQGYRLTLVRAEPQTSDEPG 1234
 QY 1145 SSSYVYLKKEEDSVSSQSSDAGLSGDSHESDTLITDVSAISNLIRKHYSEARVEDIGHEL 1204
 Db 1235 MASSSGGPPQLSNCSNEMO-----VSOFIRKHYASLSLVSDTSTEL 1274
 QY 1205 TYVLPYEAKGCAVELFHEIDRLSDGLSSYGISSETTLEEIFLKVAEE----- 1254
 Db 1275 SYLIPSEAVKKGAFERLPQGLEHSIDLALHLSFGIMDTTLEEVFLKYSEEDQSLNSEAD 1334
 QY 1255 -----SGYDA-ETSDGTL-----PARNRRA-FGKOS 1280
 Db 1335 VKESRKDALPGAEGITLAVESQAGNLARCSELAQASIASVSASVSGARAGGAGAYTGYG 1394
 QY 1281 CLRPTEDDAADPNDSIDDPRESRETDLSGMDKGSYQVQKMLTQOQFVALLKRLILIA 1340
 Db 1395 DYRPLF-DNLODP--SVSLQDAEKEALARY-GQSKRLBEMMLKMQFGLLYKRRHCA 1450
 QY 1341 RRSRKGFAQIVLPAVFCIALVFLIAPPKGYPSLELOPMYNEQY-----FVS 1392
 Db 1451 RNSKALCSQLLPAFFVCVAMTVALSVBEIGDLPLVLSPSQYH-NYQGRGNFIPIYAN 1509
 QY 1393 NDAPF-----DTGTELLNALTKDPGFGTGM----- 1419
 Db 1510 BEERREYRLRSLPDASPDQVSTFLRPSGVGATCYLKSFANGSLGPMNLSSGESRLAAR 1569
 QY 1420 -----EGNPI-----PD-----TPQAGEEEMTTAP 1440
 Db 1570 FFDSCLESFTQGLPLSNFVPPSPAPASDPLSPDEBSLLAMNTSLPRTAGPETWMAP 1629
 QY 1441 -VPQITMDLFQNGNMTWQNPSPACQSSDKIKKMLPYCPGAGGLPPQKQNTADILLQD 1499
 Db 1630 SLPLRVLHEPVR-----CTCSAQGTGFS---CPSSVG-HPQOMVYVYGDILTD 1673

```

QY 1500 LTRGRTSDLVKTYOIIIAKSLKNIWNEFRYGGESGVSNTQALPPSQEVNDAIKOMK 1559
DB 1674 IIGHNVEYLITSDRF-----RLHRYGALTEG--NIQKSIPIAIGTRPLMWR 1720
QY 1560 KHLKLAKDSSAARFLNSIGRFMTGIDTRNNKVMFNKNGMHAISSFLVANNATIRALQ 1619
DB 1721 K-----IAVRVAVQVLTNNKGHSMPTLYLNSLNAITIRALP 1757
QY 1620 KGE-NESHYGITAFNHPILNLTQKQLEVALMTTSVDVLSICIVFAMSPVASFVFLIQ 1678
DB 1758 KSKGNPAAYGITVTHNPMKTSASIS-LDYLLQGDVVAIFITVAMSFVASFVFLVLA 1816
QY 1679 EAVSAKHLQFTSGKPVYLYLNSVWDMCNVVAALVITITFCFOQKSVYSTNLFPVL 1738
DB 1817 EKSTPAKHLQFVPCNPVLYMLANVMDLNLVPAATCCIIILFFEDLPAYSPTNPAV 1876
QY 1739 ALLLLGMSITPLMPASFEVKIPSTAYVVLTSVNLFTGINSVATPVELEFT--DNKLN 1797
DB 1877 LSLFLILGMSITPLMPASFWEVSSAVFLVITNLFTGATATVATFLQLEFHKDKL 1936
QY 1798 NINDIKSVELLFPHCLGRLIDMKQAMADALERGE--NRFVSPLSMDLVGNLFAM 1856
DB 1937 VVNSYLSKSCFLFPMYNGHGLMEIAYNEYAKIGQFDKMSPEEMDIIVRGIVAM 1936
QY 1857 AVEGVAFELITLYRREFIRPRVNAKLSPLNDEDEVRRERQIILDGCGNDILEIKE 1916
DB 1997 TVEGVGFELITMCOYNEFLRQPLRVPSTKPEVD--DVVASERQVLRGDADNDMVKLEN 2055
QY 1917 LTKIYRRK---RKPVADRIQVIGIPPEGCFGLDVGAGKSSTFKMLTGDTVTGDAFLN 1973
DB 2056 LTKYKSRKIGRIILAVDRICLGVRFEGCFGLDVGAGKSTFKMLTGDESTTGEARVN 2115
QY 1974 KNSIISNHEVQNMNGYQOPQDAILTLLTGEHEFFALLGVPEKGVKGEWAIRKLQ 2033
DB 2116 GHSVTKDLQVQOSLGCPQDPAFLDELTAHEHQLQYTRLGIPMKDAQVVALEKLE 2175
QY 2034 LVKYEKAYAGNYSGNKPKLSTAMALLIGPPVFLDEPTTGMPPKARFLNCAISVYKE 2093
DB 2176 LTKCADKPAAGSYSGNKRKLSTALALIGYPAFLDEPTTGMPPKARFLNLLDLIKT 2235
QY 2094 GRSVYLTSHSMEECALCTRAIYVNGFRCLGSYQHLKNRFGDGYITIVRIAGSNPDK 2153
DB 2236 GRSVYLTSHSMEECALCTRAIYVNGFRCLGSYQHLKNRFGDGYITIVRIAGSNPDK 2294
QY 2154 PVODEFGIAPFGSYLKEKRRMLQOLPSSUSLARIPSLISQSKRRILHEDYSVQDTL 2213
DB 2295 DVYREFNENFEDAMKEKHHKVOYLKSEHISLAQVFSKMEHVVGLIEDYSVSOTTL 2354
QY 2214 DOVFVNAKQSD 2227
DB 2355 DNVFVNAKQSDN 2368

```

```

RESULT 9
Q9HC28 PRELIMINARY; PRT; 2436 AA.
AC Q9HC28;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE ATP-binding cassette sub-family A member 2 (ABC transporter
  ABCA2).
GN ABCA2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Vulevic B., Chen Z., Davis W. Jr., Walsh E.S., Tew K.D.;
RT "Cloning and characterization of human ABCA2."

```

```

RL Submitted (Aug-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX PubMed-11178988;
RA Kaminski W.E., Piehler A., Pullmann K., Porsch-Ozcurtun M., Duong C.,
RA Bared G.W., Buchler C., Schmitt G.;
RT "Complete Coding Sequence, Promoter Region, and Genomic Structure of
  the Human ABCA2 Gene and Evidence for Sterol-Dependent Regulation in
  Macrophages."
RL Biochem. Biophys. Res. Commun. 281:249-258(2001).
DR EMBL: AF178941; AAC09372.1; -.
DR EMBL: AF327657; AAK1434.1; -.
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR003439; ABC_transport.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000572; Euk_Mb_oxred.
DR InterPro: IPR000566; Lipocin_cytfABP.
DR Pfam: PF00005; ABC_Tran; 2.
DR Prodom: PD000006; ABC_transport; 2.
DR SMART: SM00382; AAA; 2.
DR PROSITE: PS00211; ABC_TRANSPORTER; UNKNOWN_1.
DR PROSITE: PS00022; EGF_1; UNKNOWN_1.
DR PROSITE: PS00213; LIPOCALIN; UNKNOWN_1.
DR PROSITE: PS00559; MOLYBDOPTERIN_EUK; UNKNOWN_1.
KW ATP-binding.
SQ
SEQUENCE 2436 AA; 26995 MW; E044A3AF14EA25D1 CRC64;

Query Match 35.9%; Score 4240.5; DB 4; Length 2436;
Best local similarity 39.8%; Pred No. 2e-269;
Matches 999; Conservative 347; Mismatches 729; Indels 435; Gaps 61;

QY 6 QLRLLKNTLFRRTQTCOLLLEVAMPLEFILISVRLSYRPPYDQHECPNKA-MPSA 64
DB 6 QLRLLKNTLFRRTQTCOLLLEVAMPLEFILISVRLSYRPPYDQHECPNKA-MPSA 65
QY 65 GTPVVOGIIICNANNPCFRPTGGEAPGVGNPNKSIYARLSQARRL---LYSQMDT 120
DB 66 GILPVMQSLCPDQREBFGL-----QYANSTYQLLERIDRYVEECSNLPDAP 115
QY 121 SMKDMKVLFTLQILKSSNLKLOFLVDNETFSG-----FLYHNLSTP 165
DB 116 SLGSELEALR--QHLBALACPGTSGSHDRSTVSSFLSDVARNQDELMTLQNLSP 173
QY 166 KSTVDMKLADY---ILKVPLOGLHLTS-LCNGS-----KSEEMI--- 204
DB 174 NSTAQLAARVDPPEVYHLDFGPSSALDSQSLKRGQEPWSRLGNPLFRMEELLAPA 233
QY 205 ---QL---GDQVSELCGLPKKELAAE-----RVLRSNDI 235
DB 234 LLEQLCTPGSGELGRLILVPESSQKALQGYRDVAVCSQAARARRFSGLSNELRNQDV 293
QY 236 LKPIELTINSTPFSKELAE--TKTLIHSLGTLAQLEFSKRSWDMKQEVMTLVNVS 293
DB 294 AK-VSQQLDLPAPNGSDSSPOAPPPRRLQALLGDLLD-----AQKVLQDVVLS 341
QY 294 SSSSTQIYQAVSRICVGHGEGGLIKSLNMYEDNNYKALRGNGTEEDAEFTYNSTTP 353
DB 342 ALALLLPGACIGRTGPPASAGGAAN-----GTGAGAVMGPNTAFEGAPSAALATP 396
QY 354 YCNLMKLNLESSPLRIKALKPLLVG----- 381
DB 397 ---DTLQGGCSAFVQ--LWAGLQPIICGNNRTIEPEALRGNMSSLGFTSKQORNLGLLY 451
QY 382 -----KILYTPDTPATROVAEYVNAKTFQELAVYHDLGEMMEELSPKIKTFMENSQEMD 434
DB 452 HLMTSNPKLILYAPAGSEVDVRLKANETFAFGVNTHYAOWNLINSAEIRSFLEQGRLOQ 511
QY 435 LVKMLDLSRDNDHFEQOQDGLDWTADQIVAFIAKHPEDVOSNQSQVYTWREAFNEN-- 492
DB 512 HLKRWL-----QYVAELRIHPE---ALNISLDELPPALRQDQFS 547
QY 493 -----QAITFTIS-----RFMECVNLKLIPEIATEVWLINKSMF--LLDERKFWAG 535

```

Db 548 LPSGALLQDLDIDNMACGWIQFMSKVSVDIFKGFPEDESVINYTLNOAVQDNVFEAS 607
 QY 536 IYFTITGSGIELPHVKKYKTRMDIDNVERTNKKIDGWDGCPADPREDMRYWGGFAY 595
 Db 608 VIFQTRKSGS--LPEHVKYKIRONSSEFEKTEINERRAYMRGPNTG--GRFYLYCGVW 662
 QY 596 LODVVEQAIIRVLCTE--KKTGYVQMQPQPCYVDIDFLRWSSRMPLFLMIATYSVAV 654
 Db 663 IODMMERAIIDTFVGHADVREGSYQMPFPCTYTRDDFLVIEHMPICMISWYSVAM 722
 QY 655 IIKGIYVEKARKETEMRIMGLDINSIMFSWISLPLVLSAGLLVYLKGLMLPYSD 714
 Db 723 TIOHIAEKEHKLKEMTKMGLNNHVMWFITGFVOLSISVTLATLTKYCGVLMHSH 782
 QY 715 PSVIVFVLSFVAVVITLLOCFILSTLFSRANLAACGGIITVLYLRYLC-----VAAQD 769
 Db 783 VVIIMFLAVVAVATIMCFVLVSLYSKAKLASACGGIITVLYLRYLYVAIREVADK 842
 QY 770 YVFLTKIFASILSPVAFGCEYFALFEEDGIGVQMDNLESPVEEDGFNLITISML 829
 Db 843 ITAFI-KCIASIMSTTATGCKSKYFALYEVAGVGIQMHTFQSPVEGDFFNLLAVTLM 901
 QY 830 FDFLYGVMTWYIEAVFPQYQIRPMYFPCTKSYWGE--ESDEKSHGSGNOKRMS-- 884
 Db 902 VDAVYVGLTWYIEAVHFGMYGLPRPMYFPLOKSYWLSGSGTEAMWGMWPARFPRLSV 961
 QY 885 ---ETC-----MEEPHKLGVSTONLVKYRQOMKAVDGLANFEGQIT 929
 Db 962 EDDQACAMBSRFEETRGMEEPHPLPLVVCYDKLTKYKDKDKLKLKLSLNTLENOV 1021
 QY 930 SFLGNHAGKTTMGLTGLFPPTSGTAYILKDIRSESTROMIGVCPQHNVLFDMLT 989
 Db 1022 SFLGNHAGKTTMGLTGLFPPTSGTAYILKDIRSESTROMIGVCPQHNVLFDMLT 1081
 QY 990 VEBHIMFYARKLGSEKHYKAEEMOALDVLPSKSLKSKTQSLSGQMKRISVALVYG 1049
 Db 1082 VEBHIMFYARKLGSEKHYKAEEMOALDVLPSKSLKSKTQSLSGQMKRISVALVYG 1140
 QY 1050 GSKVVLIDPPTGVPYSRGIMWELLKRYROGRTIILSHHMDADVIGDRATISHGKL 1109
 Db 1141 GSRATILDEPTGVPYARARATMDLITKPKPRTIILSHHMDADVIGDRATISHGKL 1200
 QY 1110 CCVGSFLKNOLGTYLTLVKKWDESSLSSCRNSSSTVSLKREDVSQSSDAGLGS 1169
 Db 1201 KCCGSPFLKGTGYDGYRLTLVKYRAEPG-----GQCPGLAS 1238
 QY 1170 DHESDTLIDVS--AISNLIRKHAVFARLVEDIGHETLVLYEAKKBCAFELFHEIDD 1227
 Db 1239 SPFGAPLSSGSELOVOPFIRKHVASCLDSTSTELSYLPSSEAKKGAFFRLFOHLER 1298
 QY 1228 RLSDIGISSYGISETLEIFLKVAAE-----SGVDA-ETSDGTLR----- 1267
 Db 1299 SLDALHLSFGMLDITLLEVFLKVSFEEDOSLENSADVYESKRDVLPKAGEAGSGBGHAG 1358
 QY 1268 -----ARRNRA-FGDKOSCLRPTEDDAADPNDSDIDPESR 1303
 Db 1359 NLARCELTQSOASISQASVSGSARGDEAGYTDVYGDYRPLF--DWPQCPD--NVSIGLV 1415
 QY 1304 EYDILSGMGKSGYQKKGKTLQOQFVALLMKRLIARSKRGFAQIYLVPAVPCIALY 1363
 Db 1416 EAAALSRV--GQGRKLDGWLKROFHGLLVKRFHCARNSKALPSQILLPFAFVCVANT 1474
 QY 1364 FSLIVPFGKPSLELOPMWNYDOYV-----FVSNAPF-----DTGSLLELNL 1408
 Db 1475 VALSVETIGDLPVLVSPSOYH--NTQOPGNFLPYANERREYRLRLSDASPOQLVSRF 1533
 QY 1409 TKDPGFGTRCM-----EGNPI----- 1424
 Db 1534 RLPSGVGATCVLKSPANGSLGPTLNLSCGESRLARFDSMCLSEFTGLPLSNFVPP 1593
 QY 1425 -----PD-----TPCQAGEEWITAP--VQOTIMDLFONONMNMONSPAC 1463
 Db 1594 PSPSPDSPASPEDILOAMNVSILPPTAGPBMNTSAPSLPRLVREYV-----C 1641

QY 1464 QSSDKTKKMLPYCPGAGGLPPQKRONADILLODITGRNISDYLVKTYVQIIAKSLKN 1523
 Db 1642 TCSAOGYGSF--CPSSYVG-HPQMRVVTGDLITDTGHNHSEVILFSDNF----- 1690
 QY 1524 KIWNERFYGFSLLGVSNQALPPSOEVNDALIKOMKHLKLAKDSSADRLNSLGRPMTG 1583
 Db 1691 ----RLHRYGATIFG--NVLSISAPSGTRAPPMVK----- 1721
 QY 1584 LDRNNVYKVFNNKGMHAISSPLNVLNALLRANLQKCE--NPSHGITAFNHPNLTKQO 1642
 Db 1722 IAVRAAOVYNNKNGHSMPTYLNSLNNALLRANLKRSGKNPAAYGITYVNHMPKTSAS 1781
 QY 1643 LSEVALMTTSVDVLSICVIFAMSFVPASFPVFLIOERVSKAKHLOFISGVKPIVWLSN 1702
 Db 1782 LS-LDYILODGTVDVIAFLIIVAMSFPASFPVFLVLAKESTKAKHLOFVSGCNPIIYWLN 1840
 QY 1703 FVWDMCNVVPATLVIIIFLCFOOKSYVSTNL-PVALLLLYLGMSITPLMPASVFEXI 1762
 Db 1841 YVMDLVNVLVAPCCVILFVFDLPATSPTNFPAVLSLEFLYGWSITPLMPASFVFEV 1900
 QY 1763 PSTAVVVLTVSNLFIGINGSVATFVLELT-DKLNINDILKSVFLPHPCGLGLD 1821
 Db 1901 PSSAYVFLVILNLEIGITAVVAFLLQLEFHDKLVKNYSYLSCLDIPVNLGHLME 1960
 QY 1822 MVKQAMADALERPGE-NRPVSLMDLVGRNLFMAVEGVFELTVLQYRFFTRPP 1880
 Db 1961 MAYNEYINEYAKIGQFDKKSPPFENDIYTRGLVMAVEGVFELTLMQYFLRPPQ 2020
 QY 1881 VNAKLSPLNDEDEDVREKORILLDGGGNDILEIKELTKYRKK--RKPADRICVGP 1937
 Db 2021 MPVSTKFEVD-DGVVASEBQFVRLERGDADNDMVKILENTKYKSKIKGLIADVRLCYVR 2079
 QY 1938 PGECFGLGVNAGKASTFEMKLTGDTVTYTRGDAPFLNKSILSNHNEVHOMGVCPOFDAT 1997
 Db 2080 PGECFGLGVNAGKASTFEMKLTGDESTTGGEAVNCHSVLKELOVQOSLGCPQCDAL 2139
 QY 1998 TELLTGREHEFFALRKEVEKYGEMAIRKILVYKGYKACNYSGNKRKLSTAM 2057
 Db 2140 FDELTARHLOLYRLKLGISWKDEARVYKMALEKLELTAKYADPAGYSGNKRKLSTAI 2199
 QY 2058 ALIGPPVVEIDETTTGMDPKARFLNMCALSVKBEBSVYVLSHSEBECALCTRATM 2117
 Db 2200 ALIGYPAFILFDEPTTGMDPKARFLNMLLIDLTIKTRSVLVLSHSEBECALCTRATM 2259
 QY 2118 VNGRPRGCGVOHLKNGRGGDTTVVRJAGSNPLKPYOFGFGLAFGSVLKEKRNMLQ 2177
 Db 2260 VNGRKLCTGSIQHLKNGRGGDTMITVR-TKSSQSVKDYVRFNRPFEALIKERHNTKVQ 2318
 QY 2178 YQLPSSLSLARIFSILSQSKRRLHIEDYSVQTTLDQVFNFAKQSD 2227
 Db 2319 YQLKSEHSLAQSVMQVSGVIGIEDYSVQTTLDVFNFAKQSDN 2368
 RESULT 10
 Q9UP00 PRELIMINARY: PRT: 1529 AA.
 ID Q9UP00: Q9NSL2:
 AC 01-MAY-2000 (TReMBLrel. 13, Created)
 DT 01-MAY-2000 (TReMBLrel. 13, last annotation update)
 DT 01-JUN-2002 (TReMBLrel. 21, last annotation update)
 DE K1A1062 protein (Fragment).
 GN K1A1062 OR DKFZP547P193.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRAIN;
 RX MEDLINE=99397452; PubMed=10470851;
 RA Kikuno R., Nagase T., Ishikawa K., Hirosewa M., Miyajima N.,
 Tanaka A., Kotani H., Momura N., Ohara O.

"Prediction of the coding sequences of unidentified human genes. XIV. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro".
 RL DNA Res. 6:197-205(1999).
 RM (2)
 RN SEQUENCE OF 1157-1529 FROM N.A.
 RC TISSUE=BRIN;
 RA Bloecher H., Boecher M., Brandt P., Mewes H.W., Well B., Wiemann S.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB028985; BAB3014.1; .
 DR EMBL; AL162060; CAB82398.1; .
 DR InterPro; IPR003439; ABC_transporter.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR000566; Lipocalin_cyFABP.
 DR Pfam; PF00005; ABC_tran; 2.
 DR ProDom; PD00006; ABC_transporter; 2.
 DR PROSITE; PS00211; ABC_TRANSPORTER; UNKNOWN_1.
 DR PROSITE; PS00223; EGF_1; UNKNOWN_1.
 DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
 KW Hypothetical protein.
 FT NON_TER
 FT SEQUENCE 1 1529 AA: 170059 MW: 362655F5E4AACAA CRC64:
 Query Match 28.4%; Score 3345.5; DB 4; Length 1529;
 Best Local Similarity 46.7%; Pred. No. 7.8e-211;
 Matches 721; Conservative 207; Mismatches 381; Indels 235; Gaps 33:
 QY 836 GVMVPIEAVPVGQIGIRPWPCTSKSYWCF--ESDESHSGSNOKRNS-----ETC 887
 DB 1 GILMTYIAVHPGMYGILFRPWFPLQKSYWLSGRTSEWMSWMPARTPRISVWEEDQAC 60
 QY 888 -----MEEPHTHLKLVSIONLVKVVYRDGKAVAVDGLALNFEGQITSEFGN 935
 DB 61 AMESRREETNGMEEPHTHLPLVYCVDKLFLKYKDKKIALNLSLUNQVVSFEGHN 120
 QY 936 GAGGTTMSILTLGFPPTSGTAYILGKDISEMSTIQNLGVCPQHNVLFDMLTVEEHIV 995
 DB 121 GAGGTTMSILTLGFPPTSGTAYILGKDISEMSTIQNLGVCPQHNVLFDMLTVEEHIV 180
 QY 996 FYARLKGISEKHVAKEMOMALDVLPSKSLKSTQSLGSMOKRLSVALAFVGSVVI 1055
 DB 181 FYSLKSMAGOEIRREMDKMTEDLEL-SNKRHSILVQTLSCGMKRLSVALAIFVGSVVI 239
 QY 1056 LDEPTAGVDPYSRGIVWELLIKYRQGTIIILSHHMDADVLDRLAIIISHGLKCVGSS 1115
 DB 240 LDEPTAGVDPYARRAIIWDLILKYKPGRTIILSHHMDADVLDRLAIIISHGLKCVGSS 299
 QY 1116 LFLKNQIGTGYLLFLVKKDVSSLSGCRNSSSTVSYLAKEDVSQSSDAGLGSCHESDT 1175
 DB 300 LFLGTYGDTGRLTLVLRPAEPG-----GPQEPGLASSPPGRA 337
 QY 1176 LTIDVS--AISNLTIRKHSARLEVEDIGHELTIVLPEYAKEGAFVLEFIEDRLSDIG 1233
 DB 338 PLSSCELOYSQFIRKHAVSCILVSDTSTLSTILPSEAKKAFELFQHLERSLDALH 397
 QY 1234 ISSIGISETILELFLKVAEE-----SGVDA-ETSDGTLF----- 1267
 DB 398 LSSFGIMDTLLEVLKVESEDSLENSADYKESRKDVLPGAEGPASCGHAGNLARCS 457
 QY 1268 -----ARRNRRA-FGDKQSCLRPFTEDDAADPNSSDIDPESREIDLS 1309
 DB 458 ELTQSQASLOSASSVSGARDEGAGTYDYGDRPLF-DMPQDPD--NVSLOVEAEALS 514
 QY 1310 GMDGKSGYVKGKMLTQOQFVALLMKRLLIARRSRKGFQAIPLPAVFQIAIVFSLIYP 1369
 DB 515 RV-GQGSRLKIDGMLKRVQFHHGLLVKRFHCARRNSKALFSQILLPFFVCVAMTVALSP 573
 QY 1370 PFGKYSLELQPMKMYNBOYT-----FVSNDAPE-----DTGLELNLALITDQPF 1414
 DB 574 ETIDLPPLVISPQYH-NYTOPRGNFIPYANBERREYRLRLSPDASPOQVASFRLPSGV 632
 QY 1415 GTRCM-----EGNFI----- 1424

DB 633 GATCVLKSPPANGSLGPTLINISGESERLLAARFDSKLESTQGLPLSNFVPPSPAPS 692
 QY 1425 -----PD-----TPCAGEEWTAP-VPQITMDLEFGONGNMTMQSPACCCSDK 1469
 DB 693 DSPASPEDDQLQANVSLPPTAGEMWTSAPSLPRLVREVR-----CTCSAG 740
 QY 1470 IKKMLVPCPPGAGGLPPPOKQNTADILDLGRNISDYLVKTYVOIITAKSLKRIWNE 1529
 DB 741 TGES-----CPSSVVG-HPPMRVYTGDIITDIGHNSEYLLFTSDRF-----RL 785
 QY 1530 FRYGCSGLGVSNTQALPPPOEVDALIKOMKHLKLAKDSSADRPLNSLGRFMTGIDTRN 1589
 DB 786 HRRGATTFG--NVLKSTIPASFGTRAPMYRK-----IAVRA 820
 QY 1590 VKVFNKNGHAISSFLNINNAIIRANLQGE-NPSHYGITAFAHPLNLTQKQLEVAL 1648
 DB 821 AQVFYNNKGHSMPTYLNSLNNALIRANLPKSKGNPAAGITVTNHPMKTSASLS-LDY 879
 QY 1649 MTSYDVIVSICVIFAMSVVPASVYFVLEQEVSAKHLQFISGVKPYIYIWSNFVDMC 1708
 DB 880 LLOGTVVTAIFLIYAMSVFVPAFVFLVAKSTAKHLQFVSGCNPIIYWLANYWDM 939
 QY 1709 NVVPATLVIIIFICQOQSVYSTNLPLYLALLLLYGMSTPLMYPASVFKIPSTAYV 1768
 DB 940 NYLVPATCCVITLFPVDLPATYSPINFPVLSLFLLYGMSTIPIMYPASFVFEVSSAYV 999
 QY 1769 VLTSVLFTGINGSVATFVLELT--DNKLNINILKSVLELIPHFCLGRGLIDWKNOA 1827
 DB 1000 FLVLINLFLGITATVATFLLQFEHDKDLKLVNYSKSCFLFEPYNIIGHGIMEVANEY 1059
 QY 1828 MADALERFGE-NRFVSPISMDLVGNLFPMAVEGVVFETIYLYOYRFRIRPVNALS 1886
 DB 1060 INEYVAKIGQFQKMSPFEMDLYTGLVAMAVEGVVGLITIMCOYINLRRORMPVSTK 1119
 QY 1887 PLNDEDEDVRRERQRLIDGCGNDILEIKELTKIYRRK---RKPVADRICVGPPEGCEG 1943
 DB 1120 PVED--DVDASEERORVLRDADNDVMIENLTKVYKSKRIGILLAVDLCLGVRGEFG 1178
 QY 1944 LIGVNGAGKSTFEMKLTGDTYTRGDAPLNKSLISNHEVQNNNGYCPQDQATITELLTG 2003
 DB 1179 LIGVNGAGKSTFEMKLTGDTYTRGDAPLNKSLISNHEVQNNNGYCPQDQATITELLTG 1238
 QY 2004 REHVEFFALLRGVPEKEVGKEMAIRKILGYKGEKYAGVNGGNKRLSTAMALGGP 2063
 DB 1239 RHHQLQYTLRLKRISSKDEKARVYKALLEKLELTKYADKAGITVSGGNKRLSTALALGTP 1298
 QY 2064 PVEFLDEPTTGMDPKARRPLMNCALSVYKEGSRVYLTSHSMECEALCTRAIMVNGFR 2123
 DB 1299 AFIFLDEPTTGMDPKARRPLMNLIIIDLKIGRSVYLTSHSMECEALCTRLAIIVNGRLR 1358
 QY 2124 CLGSVOHLKNRFGDGYTYVRIAGSNPDLKPYOPEFGIARFGSVLYKEHNRMLLOYLPSS 2183
 DB 1359 CLGSTOHLKNRFGDGYTYVRIAGSNPDLKPYOPEFGIARFGSVLYKEHNRMLLOYLPSS 1417
 QY 2184 LSLAIRFSILSSKKRLHIEDYSQTTLDQVFNFAKDDSD 2227
 DB 1418 HTSLAQVFSKMQVSGVIGIEDYSQTTLDQVFNFAKDDSDN 1461
 RESULT 11
 096J73
 ID 096J73 PRELIMINARY; PRT: 2277 AA.
 AC 096J73;
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE ATP-binding cassette transporter family A member 12.
 GN ABCA12..
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxId=9606;
 RN [1]

[illegible]

Db	869	FFPPIVLVTENELSVYLVKVMFSLPTFAVSASQYLAREEEOGIGLOEMNMTSPVODD	928
Qy	818	GENLTITSIS-----MMLDFELFICVMTWITIEAVEPQOYGI PRPMXFPCTKSYW--FG--E	868
Db	929	-----TTSFGMLCCILILADSEIYFLIWMYRVNFPCTYGMAAFWFPIITLTPWKEPFGCAE	984
Qy	869	ESDEKS-----HPSNOKRMEIOMEDEEPHILKUSYIOMLYVVRDQKVA	915
Db	985	VKPEKSNGLMEFTNIMQNTNPSASPEYMTSSNIEDEPPDOLYVAGLHVTKIY--GSKVA	1042
Qy	916	VDGLALNFEQOITSPFLGHNGACKTTMTSLGLEPPPTSGAYIILGKRIRESEMSTIQNL	975
Db	1043	VDNLTNLFYEGHITSLIGPBGACKTTISMLTGLFGASAGTIFYVGKDIKIDLTPTVKNM	1102
Qy	976	GVCPQHNVLFDMLTVEEHILWFYARLK--GLSEKHVKAEMQONALDVGLPSSKLKSKTSOL	1033
Db	1103	GVCMQHODVLSYLTTEKHELLLSYGIKVPWMTKQJHEEVKRTLKDTGLYSHRK--RVGTL	1161
Qy	1034	SGGMOKLSVALAFVGSQKVVILDEPTAGVDYPSRGCIWELLKLRGRTIILSTHME	1093
Db	1162	SGGMKRKISTISIALIGSRVYIILDEPSTGVDCSRSIWDVYSKKTARTIILSTHME	1221
Qy	1094	ADVLGDRILAIISHGKLCVCYSSSLFELKNOLGTGYITLVKQOVESLSCRSSSTVSYLK	1153
Db	1222	AEVLSDRIALFEGGKLCGCSPPYLKEAFGDGHLHLTK-----	1261
Qy	1154	KEDVSOSSDAGIGSDHEDTLTIDVSAISMLIRKHVSEARLYMEDIGHELTYVL--PYEA	1212
Db	1262	-----KSPNANAAVCDIM--AVTMAGIOSHLPKALTKRIDGELYVLLPFPST	1307
Qy	1213	AKEGAPELFEHIDRLSDIGISSYGISETTLEIPLKVAEBSGVDAETSGGTLPARR--	1270
Db	1308	KVSGAYIILTRALDNGMDNICYGISPTVVEEYFNLTKESQKNSAMSLEHITORKIG	1367
Qy	1271	NRRAFGOKQSLREFGEDDADPNDDIDIPESREHDLISGMGOKSAYOVKMKMLTQOOV	1330
Db	1368	NSNMGIS-----TPDDLSVSSSNFTD--RDKILT---KCE--RLDFGILLKIM	1412
Qy	1331	ALLMKRLIARSRKGFPAQIVLPAFVCIATVFSILTPFEGKYPSLDELQPMWY--NEOY	1388
Db	1413	ALLIKRPHHTRNNKGLLAQVILPIYFVTTAKMLGTLRNSSYEIOISLSYIGTSEOT	1472
Qy	1389	TFVNSDABEDTGLELINALTKDPGFCFRCMEGNPIPDTPC--QAGEEMTAPVPOTIM	1446
Db	1473	AFEANYHP-----STEALVSAMDPFGIDMNCIMTNS--DLQCNKSLKEWNTSECIPTNF	1526
Qy	1447	DLPQNGMNTMNPSPACQCSDDIKTKMLPVCPPGAGGLPPPRKQONTADIIODLGRNIS	1506
Db	1527	GV-----CQSEN-----VQBCP--KFNYSPPHRRYSQOYIYNLTGHQVE	1565
Qy	1507	DYLVKTYQVILAKSLKNIKWNNEF--RXGGSFLGVSNQALPPOEVNDAIKQMKHLK	1563
Db	1566	NYLIST-----ANEFYOKRIGMSEFG-----LP	1568
Qy	1564	LAKOSSADRLNSLGRPMGTLD--TRNNVYKFPNNKGMHAISSFLVNNAILRANLOKE	1622
Db	1589	LTKDLRFD-----ITGVPAKRTLAKWYDPEGHSILPAILNSLNNLFLRVMSKXD	1639
Qy	1623	NPSHGITAHPHPLNTLKQOLSEVALMTTSSVDVLSYCIYIFAMSVTPASFPVYLQEVKS	1682
Db	1640	AARH--GIIMYSHPRPGVODQ--EDQATSSILDLVALSTIMCYSVTTSAPFYVYVREHOT	1696
Qy	1683	KAKHLOFISGVKPYIYMLSNFVDMCMCVVPAATLVILIFTCFOOKSYASTNLPVALLL	1742
Db	1697	KAKQLOHISGVAGVYCWTNFIDYMWYFLVPAFSGIIAIFKRLAPFISENNLGNVSLLL	1756
Qy	1743	LLYGSITPLMYPASFEVKIIPSTAYVVLVSNLFLGINGSVATFYULEFTDNKLN-----	1798
Db	1757	ILFGYATPSMMYLLAGLFHETGMAFIYVCVNLFGINSIVLSVYVFLSEKRPNDPTLE	1816
Qy	1799	-INDLKSVFLIPFPGCGRILIDMYKNQOMADALEREGRNFVSPLSMDLYGRULFAMA	1857
Db	1817	LISELTKLFIILFPOFGCVGLIELSODQSVIJDFLKAAYEVEYPNTEFENKIGAMFAVAL	1876


```

Db 1156 PSLLLVFQAFNVHAFTRDGHMADLLPLMLYGAIIPLMYLMSFFSAASTAYTRLT1 1215
Qy 1773 VNLFGINSVATFV---LELFTDKKINNINDLKSVPLEPHGLGGLDMVYN---- 1825
Db 1216 FILLSGI---ATFIVMTTMCIPAVKLEELSTLHVFLVLPVPHCLGMANVSFENYETR 1271
Qy 1826 -----QAMADALEFGEENREVSPLSWDL--VGRNLFMAAVGVFFLITVLIQ----- 1871
Db 1272 RCTSESLAHKCKKNIOYQESFAMSTPGYKVTSAAGGGLYTLPLLEINLMLR 1331
Qy 1872 YRFFI---RPRVNAKL---SPLNDEDEDVRRERORILDDGGQNDI---LEIKELTKTYR 1922
Db 1332 LRFELICAFRRMTLAELOKRTSVLPEDQDVAEERSKILPSPDSMLDPLLINELSKY- 1390
Qy 1923 RKRKP--AANDRICGIPRECEGGLGVNGAGSSPFKMLTGTJYTRGAPFLNKSITSN 1980
Db 1391 DOAPPLADRLSLAVQKCEPGLSFGNAGKTYFKMKLTGETTYSGAFAVGYSISSD 1450
Qy 1981 IREHONMGYCHQFDATITLTLGRHEHEFFALLRGVPEKEVKGEEMALTKGLVKYGER 2040
Db 1451 IGVKRGKMGYCHQFDALDHTGREGMLVYARLRGIPERLINACVENTLRGLLEPHANK 1510
Qy 2041 YAGNYGGKRRKISTMALIGRPVVFDEPTTGMPKARRFLMNCALSVKEGSSVLT 2100
Db 1511 LVKTYSGKRRKISTGIALIGEPVVFIDPXTGMDPVARRLLMDTVARARESGAIVIT 1570
Qy 2101 SHMECEALCTPMAIMVNGRFRCISYOHKRNFGDGTI--VVRIASNDLKPVOE 2158
Db 1571 SHMECEALCTPMAIMVNGRFRCISYOHKRNFGDGTI--VVRIASNDLKPVOE 2158
Qy 2159 FGLAFPGSVLKEKRRNMLYOLESLSLARIFSLISQSKRLHIEDYSVSGTLLDQVEV 2218
Db 1631 VDLTFPGSLIEDHQMVHHLPGCDLSWAKVFGILEKAKKGVYDYSVQSLSIEQVFL 1690
Qy 2219 NFA 2221
Db 1691 SFA 1693

```

RESULT 13

```

Qy 09VRG4 PRELIMINARY; PRT; 1713 AA.
AC 09VRG4;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE CG1718 protein.
GN CG1718.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN 111
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,
RA Arif J.F., Agbayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,
RA Ballou R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borck D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burris R.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,

```

```

RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Houston D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwa C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kisko B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lam P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mekulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclebo J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spleer E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svrtkars R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL: AE003568; AAF50837.1;-.
DR FlyBase: FBgn0031170; CG1718.
DR InterPro: IPR003439; ABC_transportr.
DR Pfam: PF00005; ABC_tran. 2.
DR Prosite: PD00006; ABC_transportr. 2.
DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
KW ATP-binding; Transport.
SQ
SEQUENCE 1713 AA; 192888 MW; 9DE2003BFB9DC1CA CRC64;
Query Match 21.1%; Score 2485.5; DB 5; Length 1713;
Best Local Similarity 32.3%; Pred. No. 4.1e-154;
Matches 622; Conservative 354; Mismatches 603; Indels 345; Gaps 49;
Qy 410 HDLESGMEELSPKLTWTEFMSQENDLVRLDLSRND--HFWEQDLGDLMTAADIYAF 467
Db 24 HKQWVITELVLPALFSLT---LVLRTVDTDEQGVRYNEQNLTDLN----- 69
Qy 468 AKHPEDQSSNG-SVYTRREAFNENFNOAIRI-----SRFMECVNKLPIATE 516
Db 70 -----LQNGGFSKFEFTLCISPVNPKLVBEAMOSLCKNKCISENATOLE----- 118
Qy 517 VMLINKSMELDERKFWGIVFTGTPGSI-----LPHWVKYKIR-----MDIDNVE 564
Db 119 -----LDYVSKNAF-AGVQFDAMANLFTENDPLPNPFHRLRPALPRATIAINMTW 170
Qy 565 RTNKRKIDGWDPGRADPPED---MRYVWGFAVLQDVVEQAIIRVLTGTEKRTGYWQ 620
Db 171 LTMRLFPITDLTGPNEDDDGGLPPGYLRBGLPLQHSLSMAVLRKSGSDPLPNVVK 230
Qy 621 QMPYPCYVDIIFLRVMSRMPFLMIAVYGAIVIKIYVEKARLKEPMRIGLDSI 680
Db 231 RYEPFAVITPDLBGMSSIMSLIILSLFTYCTYITTKYTEKQLEKVKIMGLSML 290
Qy 681 LWFSPFISLIPLVASGLVLIKLG-----NLPLSPSVVFEVSVFAVNTLQCL 735
Db 291 HMTAMVMSFIMLTISALITLVKINMSDEVAVLTHANFVALVEFLIITYVSSICFCFM 350
Qy 736 TSTLSSRANLAAGCGIIFYLYLPYLVCAVQDYGTTLKIFASLSPAVFGCEFA 795
Db 351 MATFSPKSTAAVGLTFLWIAFYIPSTINSYDLSLSSKGLMSLNTMAGFGIKITL 410
Qy 796 LPEEDGIGVOMNLPESPPEEDGFNLTSISMLPDTFLYGMVYIDAVPQGVGIRP 855
Db 411 GEEGEGEGIGVOMNFTTPVSVDTLTGLGAVMTMLVSCVYIMLICLYVQVMPGSGVRRP 470
Qy 856 WYFPGTKSYWGE--SDPKSHPGSNOKRMSCEKEEPTHLKIGVSTIONLVKYRDMK 913
Db 471 WNEPFRFRETWCGERVTGVEDIPNGHVQRODKARETPEEGKIGLQGRHLKK--RFGNK 528
Qy 914 VAVDGLALNFEGQITTSFLGHNGACKTTMTSLTGLPPTSGTAVILKQDRSEKSTIRQ 973

```

```

Db 529 MVAKGLSMNMFEDETVLLGHNGAGKTTTISMLTGMFPPTSGTALINGSDITNIEGARM 588
QY 974 NLGVCPOHNVLEFDMLEVEHIMFARLKGISEKHVAKEGOMALDVGPPSSLKTKTSOL 1033
Db 589 SLGICQHNVHLDPMKSVNHIREFSKMKLRKRAVBOEVAKLKIMEL-EDKANVAASRL 647
QY 1034 SGMGOKRLSVALAFVGGSVVILDEPTAGVDYSRRGITWELLKRYOGRTTILSTHND 1093
Db 648 SGMGMRKRLSVCCALCGDIFVVLCDDEPSSGMDPSARROLMDLLOQEKVGTILLTHFMD 707
QY 1094 ADVLGRITAIISHGKLCYSSGLFLKNOLCTGYTLVTKKDDVSSLSGRNSSSVATLK 1153
Db 708 ADVLGRITAIIMCDGLKCOGCTFPLKQYGSGLYSGVNLPGYRCYTKTCDSLKICVK 767
QY 1154 KEDSVSSSSDAGLSGHESDTLLIDVSAISMLIRKHAESVADLETSDGTLPARRRNR 1213
Db 768 RDD-----CETNEVVALINKYIPGLKPEDDIGAELSYOLPDSAS 806
QY 1214 KEGAFVEIPHEIDRLSDICISYGSSETTLEIFLKAESGVDAETSDGTLPARRRNR 1273
Db 807 AK--FEEMFGOLEOSDELHNGYGVLTSMEEVPMKVGAE-----KDN TG 850
QY 1274 AFGDKOSCLR--PTEDDAADPNDSIDIPESRETDLLSGMGKSGSYQKGMKLTQOQFV 1330
Db 851 NIKDOHEIMNGSGFRGED-----DNESVSD--GIFSEKRLLOGLQLSLNCGK 898
QY 1331 ALLMKRLIARSRKGFQAYLPAVENCIALVFSLIYPPFGKYSLELOPMWNEOYTF 1390
Db 899 AMLLKFLTYWNNKLLILQINIMPYEFVVIT--LIKTQGTFC--ELKPI----- 945
QY 1391 VANDAPEDTGTLELLNALTKDPCFTRCMENPTDITPCQAGEEWMTAPVQITMDL-- 1448
Db 946 -----TSL-----TQPLAVIYVLDKSN 963
QY 1449 FONGNMWONPSPACQSSDKIKMLPYCPGAGSLPPORKNADILLOLTGNISDY 1508
Db 964 YONG-----TG-----Y 970
QY 1509 LKTYVOIIAKSLKKIIVNENRFG-GFSLGVSNTOALPRSOEVNDAIKQMKHLKLAKD 1567
Db 971 ETANKYEDLARS-----YGSNGLDELGTQGF-----EDYILDKITQYBIN 1013
QY 1568 SSADRFNLISLGRFMTGLDTRNNKVKVFNKNGHAISSFLVYNALILKANLOKGENSHY 1627
Db 1014 S--RYL-----VAATITESKITAMLNQALHTAPLIVNANHAI--ADLFG--SSV 1059
QY 1628 GTTAFNHPILNTKQOLSEVALMTTISVDLVASIVLEAFMSFVAPSVFVLIDERYSKAHL 1687
Db 1060 KIOVTNAPLPYTTSTILSQLSTGNLGTQLASNLQFCMCFVSSIIYILFLIKERSRAKL 1119
QY 1688 OFISGVKPIYWLNSFVWDMCNVVPATLYITIFICFOOKKSVYSTNLPLVALLLLXGM 1747
Db 1120 OFVGVKWTWMLSQFICDPAASYVITALLIVITVCFOETGLSTFEGELGRVYLLLLFG 1179
QY 1748 STPLMPASAEVFKIPSTAYVVLTSVNLFIGINSVAFVVL--ELFTDNKKNINDILKS 1805
Db 1180 AVLPTIYMSLFEFREPAPGARVSLVNFEGMALFIYVYVWSSLEF-DTK--DIADILGM 1256
QY 1806 VFLIFPHCLGRGLIDMKKNOAMADALERFGENRVSPL-----SMDLV 1849
Db 1237 IFRIFPHFSLMSLNKYVTNATNACAKAGA--LPIILLCELVPOCCNLKPYFAMEEP 1293
QY 1850 G--RNLFAVEGVVFLITVLIOR-----FFIR-----PRVNAKLSPLNDEDED 1884
Db 1294 GVLPEPTVMAATGVVFLITVLEPRLLINELMFKIROLISKPPPIEGOL-----DDD 1347
QY 1895 VRREFORLLDGGGQNDI--LEIKELTKIYRRKRKPAVDRIQVIGPGEFGILLGVNGA 1950
Db 1348 VANERERILQ--MSSNELATKMLVLDRTKYYGQFM--AVNGVSLCVOVEGEFGILLGVNGA 1404
QY 1951 GKSSTFKMLTDDTIVTKDAPLANKNSILSNHEVHOMKNGICPOFDALITELLTGREHVEFP 2010

```

```

Db 1405 GKTTFKMTGDERISSGAAYVOGLSLESNMNNSIYKMGYCPQFDALDDLTGREVLRIF 1464
QY 2011 ALLGVEPEKEVGKGEAIAIRKGLVYKGEKXACGYSSGNKRKRLSTAMALIGPPVFLDE 2070
Db 1465 CMLGVQDSRLRQLESLDASFGMKHIDKOTAHYSGNKRKRLSTAIIVGSPSVIYDE 1524
QY 2071 PTCMDPKARREFLMNCALSVYKESRSVVLTSMSMEDEALCTMAIIVNGRFGICLSVQH 2130
Db 1525 PTCMDPAARQRLMNVNVCRIKDSGKSLVLSHMEDEALCTRLAIIVNGRFGICLSVQH 1584
QY 2131 LKNFPGGTYIV-----RAGS--NP-----DLKVPDFFGLAF 2163
Db 1585 LKNFSGKLLIKIKVRNLEALRQARLSGVARPDQTVPAQMSORDIDAVEKPEVETRY 1644
QY 2164 PGVLEKKEHNMLOLYPSSLSIARFISLSQKRLIHEDYSVSGOTLDOYFVNFARD 2223
Db 1645 PMSILOEYQILFTYIPILGVKMSRIFGLMESNRQOLNVEDYSVSGOTLLEIFLEFAY 1704
QY 2224 QSD 2227
Db 1705 QRED 1708

RESULT 14
ID Q8W010 PRELIMINARY; PRT; 1884 AA.
AC Q8W010;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE ATP-binding cassette transporter ABCA1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Forestier C.;
RT "AtABCA1, a full size ABC transporter homolog of the mammalian
RT cholesterol efflux regulatory protein ABCA1."
RL Submitted (Apr-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: AY032590; AK33643.1;
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR003439; ABC_transporter.
DR Pfam: PF00005; ABC_tran. 2.
DR ProDom: PD000006; ABC_transporter. 2.
DR SMART: SM00382; AAA. 2.
DR PROSITE: PS00211; ABC_TRANSPORTER; UNKNOWN_2.
KW ATP-binding.
SQ
SEQUENCE 1884 AA; 209477 MW; 93CA7AB44D5D588 CRC64;

Query Match 17.7%; Score 2092; DB 10; Length 1884;
Best local similarity 28.8%; Pred. No. 3,9e-128;
Matches 592; Conservative 307; Mismatches 601; Indels 556; Gaps 58;

QY 489 NEFNQALRITS-RFMCVNDLKL--EPITAEVWLKSNMELDERKFNAGIVGTITGGS 545
Db 59 DETNNMIDILSLKPELRVTKIFKDDILETYTISAHGVCSEVRNCSNPKIKGAVFH 118
QY 546 IELPHHVKKIRK-----DIDNVERTNKKINDGYMDPGRADPPF-----DMRY 588
Db 119 EGGRPHLFDYSIRLNHTWAAGFPNKSIMDTN--GPIYNDLEMGINTIPTMOY 169
QY 589 VMGFAYLQDYVEQALIRV-----LTGTEKKTGV-----YMOAMPYC-- 626
Db 170 SFGSFLTLQOVDVFIFASQONNDPLSHSNLSALRELPLWTLSPSVIRMWVFPPIRE 229
QY 627 YVDDIILRWKSMSPLEMLAIYSAVVIKIVYEKARLKEKTRINGLSDNIIIPSWF 686
Db 230 YVDDFQSIIVKSVMGILLYLGLFPLISRLISYVEKEQKIREGLYMGMLKDEIFHLSWF 289
QY 687 ISSLLPLVASGLLVVILKGLMLPSPDPSVVFELSVFAVVTIIQCFILSTLPSRAMLA 746

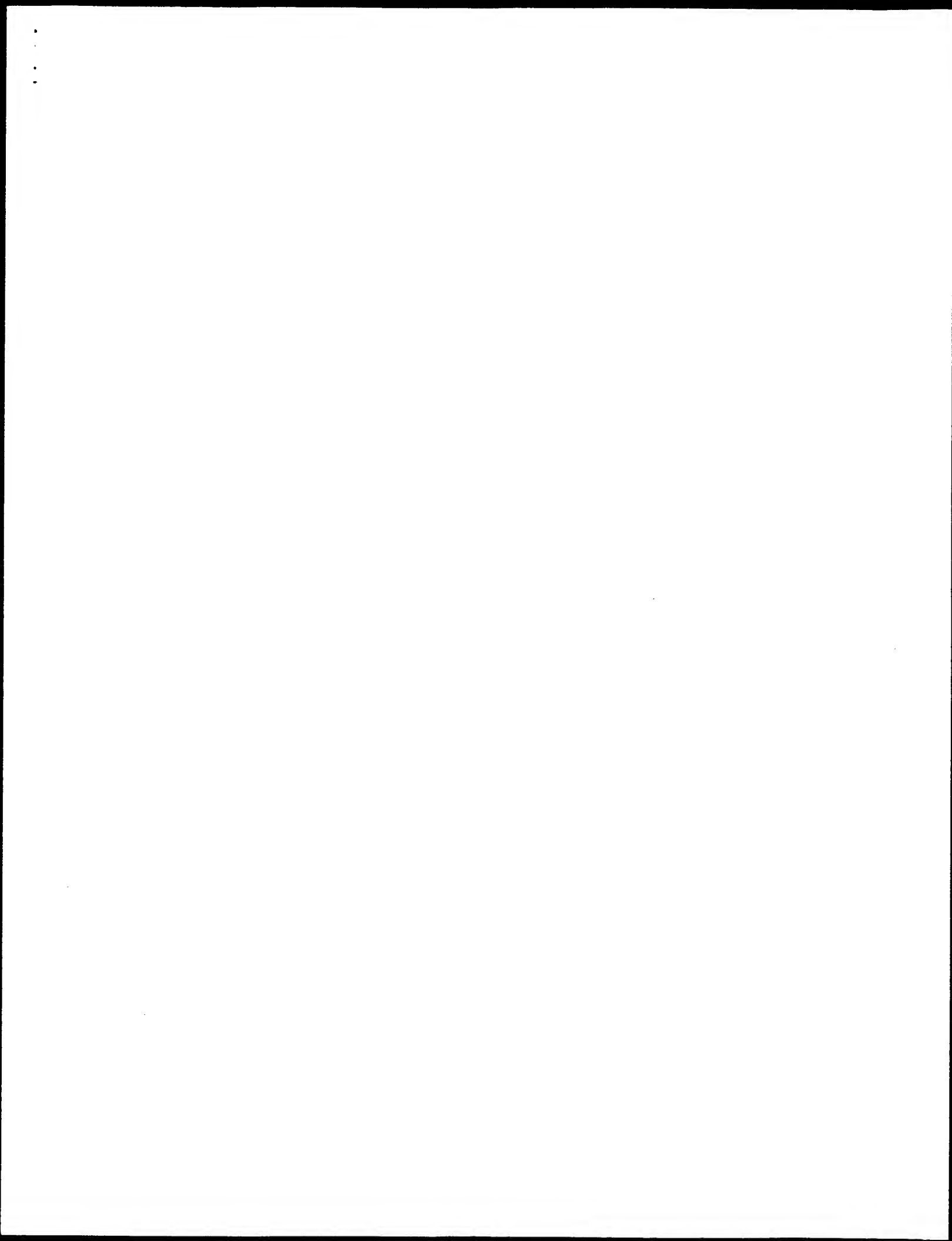
```


RA lightning J., Lloyd C., McMurtry A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
 RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierley-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sprat J., Wohlman P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans";
 RL Nature 368:32-38(1994).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Becker M., Graves T., Yeakum M.;
 RT "The sequence of C. elegans cosmid Y39D8C";
 RL Submitted (Oct-1998) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Waterston R.;
 RL Submitted (MAR-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF101313; AAC69223.1; -;
 DR InterPro: IPR003439; ABC_transporter;
 DR Pfam: PF00005; ABC_tran; 2;
 DR ProDom: PD00006; ABC_transport; 2;
 DR PROSITE: PS00211; ABC_TRANSPORTER; UNKNOMN.2;
 SQ SEQUENCE 1802 AA; 203255 MM; 933DBAEEF1AA046 CRC64;

Query Match 17.5%; Score 2060; DB 5; Length 1802;
 Best Local Similarity 27.9%; Pred. No. 4,6e-126;
 Matches 551; Conservative 338; Mismatches 630; Indels 454; Gaps 52;

QY 385 YTPDPATQVAEAVKTFQELAVFHDELGWMEELSPKWTGEMNSOEDVLRMLDSRD 444
 Db 119 YAPMDATGQIMDKIONR-----TANDL-----LNP-VAATIKGLNITAVPVLTNMT 167
 QY 445 NHFPMQDGLDWTADIVAFNAHPEDVOSSNGSVYTRPAPNETNOALRTTSREMEC 504
 Db 168 YKGF-----TTEGEMVSMQGFOS-----EC 189
 QY 505 VNINKEPIATEVWLINKEMLDERKFWAGIVTGITPGSELHHYKTKRMIDIVE 564
 Db 190 DN-----PLL-----AGIVPDDSLAKDLKNDKDFYITRLSTNH 225
 QY 565 RINKIKDGYWDGPRADPEEDMRVWG-----GRAY 595
 Db 226 RRSR-----NAGDMSYPMDSVSFAVQYSGPIMPDDNDCGSGYMQEGFMT 273
 QY 596 LODVVEQAIIRVLGTGEEK-----TGVMQOMPICYVDITFLRWKSMPLMTLANIY 650
 Db 274 VORAVDVAITELITGEDQPLPLDSYQVSRFPFGYSTKI-TEIGAFEMFVIVFESMT 332
 QY 651 SYAVILKIGIVEKEARKETMRIMGLDMSILMFSSILPLVLSAGLLVITIKGLKL 710
 Db 333 SYIIVRAVAVVEKEDRLKEMKVMKLSQFINVAHFTINAKTLFAVIVLIIIMFVAL- 391
 QY 711 PYSDPVAVFELSVAVVTLLQCFILSLFSRANLAACGGITVTLV-----GRAY 759
 Db 392 -KSDMTLMEVFLIMAYFAVUYFAFMSSEMSATSATLISVFWMLLYEWAFESSIDQT 450
 QY 760 -PYVCVAMODVVGFTLITFASLSP-VAFGGCGCYFALFPEGGIGVOMDLFESPAVED 817
 Db 451 NPYPL-----GRL-----INCLNPDIALLNGQLLAAYETQADGLMGELTPPSPDN 499
 QY 818 GENTLTSISMFLDFTLVGMVTYTEAVFGOYGP-RPWFFPCPKSYW---GEESDEK 873
 Db 500 NTFGHALLIIVDGIIMILLITWYLEAVIPGEGVPOKPMWF-VLPSPWFNNGSKYIDS 558
 QY 874 SHPGSNQKRMSTICEEPETHLKLGVSIONLVKYV-----RDGMKVAVDGL 919
 Db 559 SDOFOOIYADHVAKLEKPEPTDIPINNVNLTGYTSFFKKLPDCKGKSGEKRAVSNL 618
 QY 920 ALNFEGGITSFLHNGAGKTTMSILGLPPSGTAYILGKDIRMSMTIRONLIGCP 979
 Db 619 NLKMTPGGCVTLIGHNGAGKSTTSMITGVASPPSGSAIVNDPDIRISLPIKIREMGICP 678

QY 960 QANVLFDMLTVEEHIMYVARLKLS--EKHVKAEMQMALDVLPSKSKTSQTSQSGM 1037
 Db 679 QYNLTFCGMVMEHLEFFAKLERTWDPPEARLILARLIDF-----KADFMGALISGG 733
 QY 1038 QKSLVALAFVGGSKYVILDEPTAGVPSRKGITWELLKRYROGRITLITSHIMDAVY 1097
 Db 734 KRKSLAIALIGSEVYMLDEPTSGMDPGARHETWTLIOREKERRILLTTHMEBADLL 793
 QY 1098 GDRILAIISHGKICCVSSSLFKNOLGTGYVLTLYKKDVESLSLSCNRSSTVSYLKKDS 1157
 Db 794 GDRILAIHAGOLECCGSPMLKOYGDGYHLTY----- 827
 QY 1158 VSQSSDAGLGSDESPTLTIDVSAISNLLRKHVSEARLIVEDIGHETLYVLYPEAKEGA 1217
 Db 828 -----YDTSTPVDYSKTTDIREYIPKAVHSYIGQETVYL--SATNRP 871
 QY 1218 FVELFHEIDRLSLDGLSSSGISSETTLEELFLVAF-----EGVDATFSQGLTPAR 1269
 Db 872 FPKLKELEDHQTCGITSFGVSTTMEVEFLVGHADERYNYENGENDISE----- 925
 QY 1270 RNRRAFGDKSCLRPFTEDDADPNDSIDIPESRETDLSGMDGKSQYQKGLTQOOF 1329
 Db 926 -----MEKP--DPILOD-----RAQRYTGFTLQMOHA 953
 QY 1330 VALMKRLLIARSRKGFPAQIVLPAVEFCIALVSLVPPFGKYPSELQPMYNEQYT 1389
 Db 954 KAMFYKRAIFFEPRKWIOFLPOLVFPVAVYLVMTFSQVLPYSKE----- 997
 QY 1390 FVSNADAPEDGTLELINALIKDGFGRCMGEPIDPTPOAGEEMTAPVQOTIMDLF 1449
 Db 998 -----QDPQITSLA-----PFSQTK-KAG-----HLVSD-- 1020
 QY 1450 QNGNMTQNPSPACQSSDKIKMLPYCPGAGCLPPPOKQMTADILDLTGRNISDY 1509
 Db 1021 -SGNYVTL-----LG- -SQNLSNV 1038
 QY 1510 VKTYVOILAKSLNKIKVNFERYGGSGLYSNTQALPPSQEVNDALIKQKHILKLAKDS 1569
 Db 1039 QGTVTO-----LGVTOF-----VVDITSNVEKEFTMDOTNM 1069
 QY 1570 ADR-----FLNSLGRMTGLDTRNNYKVPFNKKMGHAISFLVNIINATLRANLQK 1620
 Db 1070 GSRFGLHVALGFVPSMFNST--VSVPILKTFEKNFGILYTPALATTFDMSMLSKQKK 1127
 QY 1621 GENBSHGTIAFNHPLNLTKQOLEVALMTTSDVLVSCVIFPMSFVASEVFLIOGR 1680
 Db 1128 -----QYSFVAVNHPLPPSPQDTLKNMNSDCAAPLILAYGLVFAVCAVAGSQFLITER 1182
 QY 1681 VSKAKHLQFISGVKPVYUWLSNFMVMCMNVVPAVLVITITFCFOOKSVVSTN-LPVL 1739
 Db 1183 KRSKHMQLSLGIRPWFMTLATITMDAPMVLIRILCFDAFIETJNTATTHDFGVALILT 1242
 QY 1740 LLLLLYGSITPAMYPASEFEKIPSTAYVVLTVSNLFINGSVAFVLELTPDNKLNMI 1799
 Db 1243 LSLFLYQWTLAPPTYMQEFEEESAPKPGMYVIMYHILTIMISIANPIISQSSSIDAGYL 1302
 QY 1800 NDILKSFLITFPFCIGRGLIDMVKQAMADALERR-----GENREVS 1842
 Db 1303 WSIIFA--WLFPTYNISQIATVTFQENENRIACKKIDCTIPMEKAVTACCGTASERLVD 1360
 QY 1843 PLMSDLVGRN-----LFMAVGVVEFLITVLYIOYRF-----IRPPVNAKSLPND 1890
 Db 1361 NVLP--VGNKKGILLVYVFLAVDGFITWVWREKNDQTKLALIRCKKADNPIMDITD 1418
 QY 1891 -----EDEDVRRR--ORILDGGQNDLLEIKELTKYIRKKRKPAVDKICVGP 1938
 Db 1419 TDKVDERVDSDSVIAEKSYVQR--ANNKKTALVSNNLVKKYGNF--AVGVMFHVNS 1474
 QY 1939 GEGEGILLGVNAGKSTSPKMLTQDITVTGDAFLKNKSLNSINHEVHQMGCCEPDAIT 1998
 Db 1475 KDFGLGVNAGKSTSPQMLTGENSEISSGDAYVNGMSVKNMWRAGANTGYCPOYDAII 1534



GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 7, 2003, 08:45:29 ; Search time 24 Seconds
(Without alignments)
2771.884 Million cell updates/sec

Title: US-09-595-526C-2

Perfect score: 11797

Sequence: 1 MACWPQLRLMKNLTPRRR.....VDVAVLTSFLQDEKESYV 2261

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued_Patents_AA.*
2: /cgn2_6/ptodata/1/1aa/5A.COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/5B.COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6A.COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/6B.COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/6C.COMB.pep.*
7: /cgn2_6/ptodata/1/1aa/6D.COMB.pep.*
8: /cgn2_6/ptodata/1/1aa/6E.COMB.pep.*
9: /cgn2_6/ptodata/1/1aa/6F.COMB.pep.*
10: /cgn2_6/ptodata/1/1aa/6G.COMB.pep.*
11: /cgn2_6/ptodata/1/1aa/6H.COMB.pep.*
12: /cgn2_6/ptodata/1/1aa/6I.COMB.pep.*
13: /cgn2_6/ptodata/1/1aa/6J.COMB.pep.*
14: /cgn2_6/ptodata/1/1aa/6K.COMB.pep.*
15: /cgn2_6/ptodata/1/1aa/6L.COMB.pep.*
16: /cgn2_6/ptodata/1/1aa/6M.COMB.pep.*
17: /cgn2_6/ptodata/1/1aa/6N.COMB.pep.*
18: /cgn2_6/ptodata/1/1aa/6O.COMB.pep.*
19: /cgn2_6/ptodata/1/1aa/6P.COMB.pep.*
20: /cgn2_6/ptodata/1/1aa/6Q.COMB.pep.*
21: /cgn2_6/ptodata/1/1aa/6R.COMB.pep.*
22: /cgn2_6/ptodata/1/1aa/6S.COMB.pep.*
23: /cgn2_6/ptodata/1/1aa/6T.COMB.pep.*
24: /cgn2_6/ptodata/1/1aa/6U.COMB.pep.*
25: /cgn2_6/ptodata/1/1aa/6V.COMB.pep.*
26: /cgn2_6/ptodata/1/1aa/6W.COMB.pep.*
27: /cgn2_6/ptodata/1/1aa/6X.COMB.pep.*
28: /cgn2_6/ptodata/1/1aa/6Y.COMB.pep.*
29: /cgn2_6/ptodata/1/1aa/6Z.COMB.pep.*
30: /cgn2_6/ptodata/1/1aa/6A.COMB.pep.*
31: /cgn2_6/ptodata/1/1aa/6B.COMB.pep.*
32: /cgn2_6/ptodata/1/1aa/6C.COMB.pep.*
33: /cgn2_6/ptodata/1/1aa/6D.COMB.pep.*
34: /cgn2_6/ptodata/1/1aa/6E.COMB.pep.*
35: /cgn2_6/ptodata/1/1aa/6F.COMB.pep.*
36: /cgn2_6/ptodata/1/1aa/6G.COMB.pep.*
37: /cgn2_6/ptodata/1/1aa/6H.COMB.pep.*
38: /cgn2_6/ptodata/1/1aa/6I.COMB.pep.*
39: /cgn2_6/ptodata/1/1aa/6J.COMB.pep.*
40: /cgn2_6/ptodata/1/1aa/6K.COMB.pep.*
41: /cgn2_6/ptodata/1/1aa/6L.COMB.pep.*
42: /cgn2_6/ptodata/1/1aa/6M.COMB.pep.*
43: /cgn2_6/ptodata/1/1aa/6N.COMB.pep.*
44: /cgn2_6/ptodata/1/1aa/6O.COMB.pep.*
45: /cgn2_6/ptodata/1/1aa/6P.COMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6931	58.8	1375	US-08-665-259-26	Sequence 26, Appl
2	6931	58.8	1375	US-08-665-259-26	Sequence 26, Appl
3	3129.5	26.5	1457	US-08-665-259-27	Sequence 27, Appl
4	3129.5	26.5	1457	US-08-665-259-27	Sequence 27, Appl
5	2642.5	22.4	1684	US-08-665-259-25	Sequence 25, Appl
6	2642.5	22.4	1684	US-08-665-259-25	Sequence 25, Appl
7	2642.5	22.4	1704	US-08-762-500-75	Sequence 75, Appl
8	340.5	2.9	1280	US-08-583-276-19	Sequence 19, Appl
9	339.5	2.9	1280	US-08-583-276-19	Sequence 19, Appl
10	337	2.9	1279	US-08-784-649A-2	Sequence 2, Appl
11	330.5	2.8	1280	US-08-752-447-2	Sequence 2, Appl
12	330.5	2.8	1280	US-08-752-447-2	Sequence 2, Appl
13	315.5	2.7	1275	US-09-134-001C-3369	Sequence 3369, Ap
14	311	2.6	1275	US-09-120-513-2	Sequence 2, Appl
15	311	2.6	1275	US-09-120-513-2	Sequence 2, Appl
16	309	2.6	1275	US-09-120-513-2	Sequence 2, Appl
17	301.5	2.5	1302	US-09-134-001C-3570	Sequence 3570, Ap
18	290	2.5	1302	US-08-232-537-2	Sequence 2, Appl
19	289.5	2.4	1349	US-08-612-734B-2	Sequence 5550, Ap
20	285.5	2.4	1349	US-08-612-734B-2	Sequence 5550, Ap
21	284	2.4	1349	US-08-612-734B-2	Sequence 12, Appl
22	283	2.4	1408	US-08-612-521-2	Sequence 3824, Ap
23	281.5	2.4	1408	US-08-612-521-2	Sequence 3824, Ap
24	276	2.3	1307	US-08-134-001C-3292	Sequence 3292, Ap
25	275	2.3	1307	US-08-134-001C-3292	Sequence 3292, Ap
26	266.5	2.3	1307	US-08-134-001C-3292	Sequence 4891, Ap
27	261	2.2	1334	US-08-996-545-2	Sequence 2, Appl

28	261	2.2	1334	US-09-328-320-2	Sequence 2, Appl
29	250	2.1	376	US-08-997-080-89	Sequence 89, Appl
30	250	2.1	376	US-08-997-080-89	Sequence 89, Appl
31	250	2.1	376	US-08-997-080-89	Sequence 89, Appl
32	250	2.1	376	US-08-997-080-89	Sequence 89, Appl
33	250	2.1	376	US-08-997-080-89	Sequence 89, Appl
34	250	2.1	376	US-08-997-080-89	Sequence 89, Appl
35	244.5	2.1	341	US-09-205-426-89	Sequence 4042, Ap
36	242	2.1	215	US-09-305-984-18	Sequence 18, Appl
37	242	2.1	215	US-09-305-984-18	Sequence 18, Appl
38	242	2.1	215	US-09-305-984-18	Sequence 18, Appl
39	240	2.0	1528	US-08-463-092B-6	Sequence 6, Appl
40	240	2.0	1528	US-08-463-092B-6	Sequence 6, Appl
41	240	2.0	1528	US-08-463-092B-6	Sequence 6, Appl
42	240	2.0	1528	US-08-463-092B-6	Sequence 6, Appl
43	240	2.0	1528	US-08-463-092B-6	Sequence 6, Appl
44	237	2.0	224	US-09-305-984-72	Sequence 72, Appl
45	237	2.0	224	US-09-305-984-72	Sequence 72, Appl

ALIGNMENTS

RESULT 1
US-08-665-259-26
Sequence 26, Application US/08665259
Patent No. 6028173
GENERAL INFORMATION:
APPLICANT: Landes, Gregory M.
APPLICANT: Burn, Timothy C.
APPLICANT: Connors, Timothy D.
APPLICANT: Dackowski, William R.
APPLICANT: Van Raay, Terence J.
APPLICANT: Klinger, Katherine W.
TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME
NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENZYME CORPORATION
STREET: One Mountain Road
CITY: Framingham
STATE: Massachusetts
COUNTRY: United States of America
ZIP: 01701
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/665,259
FILING DATE: 17-JUN-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Dugan, Deborah A.
REGISTRATION NUMBER: 37,315
REFERENCE/DOCKET NUMBER: IG5-9.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (508) 872-8400
TELEFAX: (508) 872-5415
INFORMATION FOR SEO ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 1375 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-665-259-26

Query Match 58.8%; Score 6931; DB 3; Length 1375;
Best Local Similarity 97.2%; Pred. No. 0;
Matches 1336; Conservative 20; Mismatches 19; Indels 0; Gaps 0;


```

QY 887 CMEEPHTLKLGVSIQNLVKKVYRDGMKVAVDGLALNFEGQITTSFLGHNGAGKTTMSIL 946
    |||||
Db 1 CMEEPHTLKLGVSIQNLVKKVYRDGMKVAVDGLALNFEGQITTSFLGHNGAGKTTMSIL 60
QY 947 TGLFPPTSTATVILKQDRSEMSITRONLGVCPDHNVLFDMLVEEHIFPARKLGISEK 1006
    |||||
Db 61 TGLFPPTSTATVILKQDRSEMSITRONLGVCPDHNVLFDMLVEEHIFPARKLGISEK 120
QY 1007 HYKAMEQMALDVGLPSSKLSKTSQLSGMOBKLSVALAFVGGSKVYLDEPTAGVPPY 1086
    |||||
Db 121 HYKAMEQMALDVGLPSSKLSKTSQLSGMOBKLSVALAFVGGSKVYLDEPTAGVPPY 180
QY 1067 SRRGIMELLKTRGRTIILSTHMDVADGRIATISHGKLCCVSSLEFLKNOLGTY 1136
    |||||
Db 181 SRRGIMELLKTRGRTIILSTHMDVADGRIATISHGKLCCVSSLEFLKNOLGTY 240
QY 1127 YTLVKKDVESLSSCRNSSSTVSYLKKEDVSSQSSDAGISGDSHESDTLIDVSAISNL 1186
    |||||
Db 241 YTLVKKDVESLSSCRNSSSTVSYLKKEDVSSQSSDAGISGDSHESDTLIDVSAISNL 300
QY 1187 IRKHVSEARLVEDIGHELTYVLPYPAKEGAFVELFHEIDRLSDLGISYSETTLEE 1246
    |||||
Db 301 IRKHVSEARLVEDIGHELTYVLPYPAKEGAFVELFHEIDRLSDLGISYSETTLEE 360
QY 1247 IFLKAAEESGVDAETSDGTLPARRRKRAFQKQSLRFTEDDAADPDSDIDPESRETD 1306
    |||||
Db 361 IFLKAAEESGVDAETSDGTLPARRRKRAFQKQSLRFTEDDAADPDSDIDPESRETD 420
QY 1307 ILSGMDGKGSYQVKGWKLTOQOFVALLMKRLLIARRSRKGFPAQIVLPAYVCAIALVFSL 1366
    |||||
Db 421 ILSGMDGKGSYQVKGWKLTOQOFVALLMKRLLIARRSRKGFPAQIVLPAYVCAIALVFSL 480
QY 1367 IVPFGKPTSELQPMWMEYOTFVSNAPEDGTLELNLALTKDQFGTRCMEGNPIPD 1426
    |||||
Db 481 IVPFGKPTSELQPMWMEYOTFVSNAPEDGTLELNLALTKDQFGTRCMEGNPIPD 540
QY 1427 TPCQAGEEEMTAPVPTIMDLFQNGNMTMONPSPACCCSSDKIKKMLPYCPGAGGIPP 1486
    |||||
Db 541 TPCQAGEEEMTAPVPTIMDLFQNGNMTMONPSPACCCSSDKIKKMLPYCPGAGGIPP 600
QY 1487 PQRKNTADILLODLTGRINISYLVKTYOIIAKSLKKNKWNEFRYGGFSLGVSNTQALP 1546
    |||||
Db 601 PQRKNTADILLODLTGRINISYLVKTYOIIAKSLKKNKWNEFRYGGFSLGVSNTQALP 660
QY 1547 PSEOENDAIKOMKHLKLAKDSSADRFNLSGREMTGIDTRNNKQWPNNGGMAISSFL 1606
    |||||
Db 661 PSEOENDAIKOMKHLKLAKDSSADRFNLSGREMTGIDTRNNKQWPNNGGMAISSFL 720
QY 1607 NVINNAIIRANLQGENSHYGTAFNHPNLTKOOLSEVALMTTSVDVLAISICVIFAMS 1666
    |||||
Db 721 NVINNAIIRANLQGENSHYGTAFNHPNLTKOOLSEVALMTTSVDVLAISICVIFAMS 780
QY 1667 FVPASFVVELLOEVRSAKHLQFTISGVKPYTYMNSNVMDKCNVVAIVLIIIFTCFOQ 1726
    |||||
Db 781 FVPASFVVELLOEVRSAKHLQFTISGVKPYTYMNSNVMDKCNVVAIVLIIIFTCFOQ 840
QY 1727 KSYVSTNLPVIALLLLGMSITPLMPASFVEKIPSTAVVLTSVNLFINGISVATE 1786
    |||||
Db 841 KSYVSTNLPVIALLLLGMSITPLMPASFVEKIPSTAVVLTSVNLFINGISVATE 900
QY 1787 VLEFLTDKLNINDILKSVFLFPHFCGLGRLIDWYKNOAMADALERFGBNRPVPLSW 1846
    |||||
Db 901 VLEFLTDKLNINDILKSVFLFPHFCGLGRLIDWYKNOAMADALERFGBNRPVPLSW 960
QY 1847 DLVGNLFLAMAVEGVFELTVLIQYRFPILRPVNAKLSINDEDEDDVAREQRILLDG 1906
    |||||
Db 961 DLVGNLFLAMAVEGVFELTVLIQYRFPILRPVNAKLSINDEDEDDVAREQRILLDG 1020
QY 1907 GQNDILEIKELTKYTRRRKRAVDRIQGIPEGCEFGILGVNGAGSKSTFKMLTGDTTV 1966
    |||||
Db 1021 GQNDILEIKELTKYTRRRKRAVDRIQGIPEGCEFGILGVNGAGSKSTFKMLTGDTTV 1080
QY 1967 RGDALFKNKSILSNHVEHONMGYCPQFDATTELTGREHVEFFALLRGVPEKEVKGCE 2026

```

```

Db 1081 RGDALFKNKSILSNHVEHONMGYCPQFDATTELTGREHVEFFALLRGVPEKEVKGCE 1140
    |||||
QY 2027 WAIIRKGLVYKGRVAGNNSGNKRLSTAMALISGPVVFLEPTTGMDPARREFLWNC 2086
    |||||
Db 1141 WAIIRKGLVYKGRVAGNNSGNKRLSTAMALISGPVVFLEPTTGMDPARREFLWNC 1200
QY 2087 ALSVYKGRSVLITSHMECEALCTRMALMVNGFRGIGSVQHLKRRGDDGYTVVRIA 2146
    |||||
Db 1201 ALSVYKGRSVLITSHMECEALCTRMALMVNGFRGIGSVQHLKRRGDDGYTVVRIA 1260
QY 2147 GSNPDLKPVODFPGALFPGSVLKEKRRNMLQYOLPSSLSLARIFSLISQSKRLHIEDY 2206
    |||||
Db 1261 GSNPDLKPVODFPGALFPGSVLKEKRRNMLQYOLPSSLSLARIFSLISQSKRLHIEDY 1320
QY 2207 SVSOTTLDOYVFNFAKDQSDDDLKDLSLKKNQYVDVAVLTSFLODEVKKESYV 2261
    |||||
Db 1321 SVSOTTLDOYVFNFAKDQSDDDLKDLSLKKNQYVDVAVLTSFLODEVKKESYV 1375

```

RESULT 2

```

US-08-762-500-26
; Sequence 26, Application US/08762500
; Patent No. 6030806
; GENERAL INFORMATION:
; APPLICANT: Landes, Gregory M.
; APPLICANT: Burn, Timothy C.
; APPLICANT: Connors, Timothy D.
; APPLICANT: Dackowski, William R.
; APPLICANT: Van Raay, Terence J.
; APPLICANT: Klingner, Katherine W.
; TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
; TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME
; NUMBER OF SEQUENCES: 83
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENZYME CORPORATION
; STREET: One Mountain Road
; CITY: Framingham
; STATE: Massachusetts
; COUNTRY: United States of America
; ZIP: 01701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08762,500
; FILING DATE: 09-DEC-1996
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/665,259
; FILING DATE: 17-JUN-1996
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/10469
; FILING DATE: 17-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Dugan, Deborah A.
; REGISTRATION NUMBER: 37,315
; REFERENCE/DOCKET NUMBER: 195-9.3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508) 872-8400
; TELEFAX: (508) 872-5415
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1375 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-762-500-26

```

Query Match

58.8%; Score 6931; DB 3; Length 1375;

Best Local Similarity 97.2%; Pred. No. 0;
Matches 1336; Conservative 20; Mismatches 19; Indels 0; Gaps 0;

QY 887 CMEEPPTLKGVSIONLVKYYRGMKAVADGALNFEQGITSLFLGNGAGKTTTSTL 946
Db 1 CMEEPPTLKGVSIONLVKYYRGMKAVADGALNFEQGITSLFLGNGAGKTTTSTL 60
QY 947 TGLPPTSGTAVIIGKDRSEKSTIRONLGYCPQHNVLFDMLTVEEHWFAARLKJSEK 1006
Db 61 TGLPPTSGTAVIIGKDRSEKSTIRONLGYCPQHNVLFDMLTVEEHWFAARLKJSEK 120
QY 1007 HKKAEOMALDVGIPSSKIKSKTSQSGMORKISVALAFVGSKKVYIDEPPAGVDPY 1066
Db 121 HKKAEOMALDVGIPSSKIKSKTSQSGMORKISVALAFVGSKKVYIDEPPAGVDPY 180
QY 1067 SRRGIMWELLKYRGRITITLSTHHMDEADVIGDRITAIISHKCLCCVGSLELKNOLGTY 1126
Db 181 SRRGIMWELLKYRGRITITLSTHHMDEADVIGDRITAIISHKCLCCVGSLELKNOLGTY 240
QY 1127 YTLVKKVSESSLSGCRSSSTVSYLKKESSQSSDAGSGSHESDTLIDVSAISNL 1186
Db 241 YTLVKKVSESSLSGCRSSSTVSYLKKESSQSSDAGSGSHESDTLIDVSAISNL 300
QY 1187 IRKHVSEARLVEDIGHELTYVLPYPAKEGAFVELFHEIDRLSDLGISYGISSETLIEE 1246
Db 301 IRKHVSEARLVEDIGHELTYVLPYPAKEGAFVELFHEIDRLSDLGISYGISSETLIEE 360
QY 1247 IFLKAEESGVAFETSDCTIPARRRRAPGKOSCLRPETEDDADPNDSDIDPESRPTD 1306
Db 361 IFLKAEESGVAFETSDCTIPARRRRAPGKOSCLRPETEDDADPNDSDIDPESRPTD 420
QY 1307 ILSGMDGSGSYOVKMKLTQOQFVALLMKRLIARRSKGFEAQIVLPAVVCIALVESL 1366
Db 421 ILSGMDGSGSYOVKMKLTQOQFVALLMKRLIARRSKGFEAQIVLPAVVCIALVESL 480
QY 1367 IVPFGKYPSELQPMWYNEQTYEVSNDAPEDTGTLELLMALTKDPGFCIKMEGNPIPD 1426
Db 481 IVPFGKYPSELQPMWYNEQTYEVSNDAPEDTGTLELLMALTKDPGFCIKMEGNPIPD 540
QY 1427 PTCOAGEEETAPAPORTIMOLFONGNMNMNOSPACOCSSDKIKKMLPVCPPGAGLPP 1486
Db 541 PTCOAGEEETAPAPORTIMOLFONGNMNMNOSPACOCSSDKIKKMLPVCPPGAGLPP 600
QY 1487 PORKONTADILLODTGRNISDYLVKTYVOIILAKSLKNKIWNEFRYGGFSLGVSNTQALP 1546
Db 601 PORKONTADILLODTGRNISDYLVKTYVOIILAKSLKNKIWNEFRYGGFSLGVSNTQALP 660
QY 1547 PQOEYNDATIKOMKHLKAKOSSADREPLNSIGREMTGLDTRNNKVKWPNKGMHAISSEL 1606
Db 661 PQOEYNDATIKOMKHLKAKOSSADREPLNSIGREMTGLDTRNNKVKWPNKGMHAISSEL 720
QY 1607 NVINNAIILRANLOKGENPSHYGITAFNHPNLNTKOLSEVALMTTSVIVASICVIFAMS 1666
Db 721 NVINNAIILRANLOKGENPSHYGITAFNHPNLNTKOLSEVALMTTSVIVASICVIFAMS 780
QY 1667 FVPASVVFLLIOERYSKAKHLOFISGVKPVLYWLSNFWDMCNVVPATLVIIIFTCFOQ 1726
Db 781 FVPASVVFLLIOERYSKAKHLOFISGVKPVLYWLSNFWDMCNVVPATLVIIIFTCFOQ 840
QY 1727 KSYVSSTNLPVALLLLLYGMSITPLMPASVVERKIPSTAYVVLTSVNLFGINGSVATF 1786
Db 841 KSYVSSTNLPVALLLLLYGMSITPLMPASVVERKIPSTAYVVLTSVNLFGINGSVATF 900
QY 1787 VLELFNDKLNININDIKSVFLIFPHFCIGRGLIDMVKNOAADLFRGEREFVSPISM 1846
Db 901 VLELFNDKLNININDIKSVFLIFPHFCIGRGLIDMVKNOAADLFRGEREFVSPISM 960
QY 1847 DLVGRNLFMAAVEGVFLLITVLIQYREFIRPPRYNAKLSPLNDEDEVRERKORILDGG 1906
Db 961 DLVGRNLFMAAVEGVFLLITVLIQYREFIRPPRYNAKLSPLNDEDEVRERKORILDGG 1020
QY 1907 GONDIEIEIKELTKYRRKKPAVDRIICIPGECFGLGVGAGKSTFEKMLTJDDTVT 1966
Db 1907 GONDIEIEIKELTKYRRKKPAVDRIICIPGECFGLGVGAGKSTFEKMLTJDDTVT 1966

Db 1021 GONDIEIEIKELTKYRRKKPAVDRIICIPGECFGLGVGAGKSTFEKMLTJDDTVT 1080
QY 1967 RGAFLFNKNSILSNIEHVHOMGYCPOFADITELLGREHVEFPALLRGVPEKEXKAGE 2026
Db 1081 RGAFLFNKNSILSNIEHVHOMGYCPOFADITELLGREHVEFPALLRGVPEKEXKAGE 1140
QY 2027 WALIKLGVYKGEYAGN SGNKRRKLSAMALIGPPVVFIDEPTTGMDPKARFLMNC 2086
Db 1141 WALIKLGVYKGEYAGN SGNKRRKLSAMALIGPPVVFIDEPTTGMDPKARFLMNC 1200
QY 2087 ALSVKEGRSVLTSNMECEALCTRMALMVNGRRRCGLSVOHKLKNRGDCGTYIVRIA 2146
Db 1201 ALSVKEGRSVLTSNMECEALCTRMALMVNGRRRCGLSVOHKLKNRGDCGTYIVRIA 1260
QY 2147 GSNPDLKPVODFGLAFPGSVLKEKHRNMLOYOLPSSLARSIFSLSSKRLHIEDY 2206
Db 1261 GSNPDLKPVODFGLAFPGSVLKEKHRNMLOYOLPSSLARSIFSLSSKRLHIEDY 1320
QY 2207 SVSQTTLDOYFVNFPAKQSDDDLKDLSLKRNQTVVDAVAVLSFLODEKVESYV 2261
Db 1321 SVSQTTLDOYFVNFPAKQSDDDLKDLSLKRNQTVVDAVAVLSFLODEKVESYV 1375

RESULT 3
US-08-665-259-27
; Sequence 27, Application US/08665259
; Patent No. 6028173
; GENERAL INFORMATION:
; APPLICANT: Landes, Gregory M.
; APPLICANT: Burn, Timothy C.
; APPLICANT: Connors, Timothy D.
; APPLICANT: Dackowski, William R.
; APPLICANT: Van Raay, Terence J.
; APPLICANT: Klinger, Katherine W.
; TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
; TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENZYME CORPORATION
; STREET: One Mountain Road
; CITY: Framingham
; STATE: Massachusetts
; COUNTRY: United States of America
; ZIP: 01701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/665,259
; FILING DATE: 17-JUN-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Dugan, Deborah A.
; REGISTRATION NUMBER: 37,315
; REFERENCE/DOCKET NUMBER: IG5-9.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508) 872-8400
; TELEFAX: (508) 872-5415
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1457 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-665-259-27
Query Match 26.5%; Score 3129.5; DB 3; Length 1457;
Best Local Similarity 46.0%; Pred. No. 36-288;
Matches 676; Conservative 207; Mismatches 377; Indels 211; Gaps 28;

```

QY 888 MEEPTHLKGVYIGNLVKVRDGMKVAVDGLALNFYEGQITSPFGHAGACKTTMSTLT 947
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 1 MEEPTHLPIVYCVDKLTIKVYKNDKLLANKLSLNLVENVVSVFLGHGACKTTMSTLT 60
QY 948 GLEPPTSGTAVILGDIIDSEKSTIRONIGVCPQHNVLFDMLTVEHIMFWYARLGLSEKH 1007
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 61 GLEPPTSGTAVIYGHDIIDSEKSTIRONIGVCPQHNVLFDMLTVEHIMFWYARLGLSEKH 120
QY 1008 VKAEMOALVDGLPSSKLSKTSQLSGGMOKLSVALAFVGSKSVTLDEPTAGVPYS 1067
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 121 IRKETDKMIEDEL-SNKRHSILVQTLSCGMKRLSVLAFVGSRAITLDEPTAGVPYS 179
QY 1068 RRGITELLKTRKQGTITLSTHMDADVLDRIALISHGLCCVSSSLKNOIGTGY 1127
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 180 RRAIMDLIKTRKPGRTITLSTHMDADVLDRIALISHGLCCVSSSLKNOIGTGY 239
QY 1128 LTLVKDVESLSSCRNSSITVYLKEDYSOSSDAGLCSHESDITLIDVSAISNL 1187
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 240 LTLVKOPAPGTSOPEGLASSPSCGRLSSCSBPQ-----VSOFI 279
QY 1188 RKHVSARLVEDIGHELTYLPEYAKGAFVLEFHEIDRLSDGISVYSEITTEEI 1247
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 280 RKHVASLIVSDTSTELSYILPSEAVKKGAFERLEQLEHSLDLHLSPGLMDTTEEV 339
QY 1248 FLKVAEE-----SGVDA-ETSDGTLPRARRRRAFGDKQSLREPTFE----- 1287
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 340 FLKVSSEDSLENSBADVLESKRVDYLPAGEITLVAGGAGNLCARCSLGAQSASLQAS 399
QY 1288 -----DDAADPNDSIDIPESRETDLSGMDGKSGVYKGMKL 1324
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 400 VGSARGEGTGYSDGQYRFLPNLQDPD--NVSLOEAEEMALAOY-CGOSKRLDEGMWL 456
QY 1325 TQOQVALLMKRLIARSKGFFAOYLPAVFCIALVSLVPRPGKIPSELDQPMY 1384
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 457 KMROHGLVLRPHOARNSKALCSQIILPAFVCAVATVLAPEIGDIPPLVSPSQY 516
QY 1385 NBOYT-----FVSNDAPE-----DTGTELLNALTKDPEFGTRCM----- 1419
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 517 H-NTOPGRNFIPIYANEROFYRLRLSPDASPOQLVSTFRLPSVGATCYLKSANGSLG 575
QY 1420 -----EGNFI-----PD----- 1426
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 576 PMLNLSGESRLLAARFDSKLESTQGLPLSNFVPPPSAPSDSPVXPDEDSLOAMN 635
QY 1427 --TPOAGEEEMTTAP-VPOTIMLFONGMNTQONPSACOSSDKTKMLPWPFCAGG 1483
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 636 MSLPPTAPGFTWISAPSLPIRVHEPR-----CTCSAQGTGFS---CPSSVGG 680
QY 1484 LPPQKONTADILQDLTGRNISDYLVKYVOLIATKSLKNIWNEPRYSGFSIGVSNQ 1543
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 681 -HPQMRVVTGDI.LTDITGNHVSLEYLFTSDRF-----RLHRGCAITFG--NVQ 726
QY 1544 ALPPQEVNDALIKMKHKLAKDSSADRLNSIGRMTGIDTRNNVKWVFNKGMHAIS 1603
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 727 KSTPAS-----FGARVPMVRKTIARVAAQVILYNNKGYHSM 763
QY 1604 SFLVNNAILRANLQGE--NPSHYGTAFNHPNLTKOOLSEVALMTSVDVLVSQV 1662
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 764 TYLSLNNAILRANLPSKSGKPAAXITVYNNPKKISATLS-LDYLLOSTDVIAFTI 822
QY 1663 FAFMFVPAFVFLIOERSKAKHLOFTSGVKRYTWLSNFWDMCNVYPATLVIIIFT 1722
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 823 VAMFVPAFVFLVLAERSTKAKHLOFTSGCNPEYTWLVANVMDMLYVAPCCVILIF 882
QY 1723 CFQOKSVVSTNLPVLLLLLVGKSTPTLMTYPAFVFKIPSTAYVVLVSVNLFIGNS 1782
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 883 VEDLPATTSPPNPAVLSLTLGWSITPIMYPASEFWEVPSYVFLVILINFTGITAP 942
QY 1783 VAFVLELFT--DNKLTNNINDILKSVPLIFPHFCIGRLIDMVRNOAMADALEFGE--NRF 1840
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 943 VATFLQLFEDHDKLVVNSYLSKSCFLIPPNYNLGHLMMAVNEYINEYAAIGQFDKM 1002
QY 1841 VSPLSMDLVGRNLFAMAVEGVFLITVLQYRFFIIRPRVNAKLSPLNDEDEVRERO 1900

```

```

Db 1003 KSPFENDIYVIRGLVAMTVBEGVFFLITMOQYNFLQRPQRLVSTKRPVD-DYDVAASRQ 1061
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 1901 RIIDGGONDILEIKELTKYRRK---RRPAYDRICVGI-PGEGCHLIGVAGACKSSTF 1956
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 1062 RVLRGDADNDMVKIENLTKYKSRKIGRIADVRLCLGVCVGEFGGLGVAGACKTSTF 1121
QY 1957 KMLTGPTVYRGDAPFNKNSISINIEHVONNGVCPQPAITELLTGREHVEFFALLRGV 2016
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 1122 KMLTGESTTGGAFVNGSHVLKDLQVOOSLQCYCPQVPPVDELTAHEDLRLRLCT 1181
QY 2017 PEREKGVEEMALIKLGVYGEKAGYNSGGKRRKLTAMALIGPPVFLDEPTTGM 2076
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 1182 PMDEAQVVMKALEKLELTKYADKPAGTSGGKRLSTALIALIGVPAFIFLDEPTTGM 1241
QY 2077 PKARFLMNCALSVYKEGSVVITSHSMECEALCTRAIMVNGFRCLGSVOHLKNRG 2136
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 1242 PKARFLMNLIDLTKGSRVLTSHSMECEALCTRAIMVNGRLHCLGSIOLHKNRG 1301
QY 2137 DGYTIYVIRAGSNPDLKPYQDFGLAFPGSVLKERHNNLOYQLPSSLSLARIFSLSO 2196
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 1302 DGYMIVIR-TKSSQNVKDVVRFNRPFAHAGKTPRYVQYQKLSHSLAQSVMQ 1360
QY 2197 SKRRLHEDYSQTTLDQVYVNAKOSDD 2227
    ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 1361 VVGVLGIEDYSQTTLDQVYVNAKOSDN 1391

```

RESULT 4

US-08-762-500-27

Sequence 27, Application US/08762500

Patent No. 6030806

GENERAL INFORMATION:

APPLICANT: Landes, Gregory M.

APPLICANT: Burn, Timothy C.

APPLICANT: Connors, Timothy D.

APPLICANT: Dackowski, William R.

APPLICANT: Van Raay, Terence J.

APPLICANT: Klinger, Katherine W.

TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES, TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME

NUMBER OF SEQUENCES: 83

CORRESPONDENCE ADDRESS:

ADDRESSEE: GENZYME CORPORATION

STREET: One Mountain Road

CITY: Framingham

STATE: Massachusetts

COUNTRY: United States of America

ZIP: 01701

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/762,500

FILING DATE: 09-DEC-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/665,259

FILING DATE: 17-JUN-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/US96/10469

FILING DATE: 17-JUN-1996

ATTORNEY/AGENT INFORMATION:

NAME: Dugan, Deborah A.

REGISTRATION NUMBER: 37,315

REFERENCE/DOCKET NUMBER: IG5-9.3

TELECOMMUNICATION INFORMATION:

TELEPHONE: (508) 872-8400

TELEFAX: (508) 872-5415

INFORMATION FOR SEQ ID NO: 27:

SEQUENCE CHARACTERISTICS:

LENGTH: 1457 amino acids
 TYPE: amino acid
 STRANDEDNESS: not relevant
 TOPOLOGY: unknown
 MOLECULE TYPE: protein
 US-08-762-500-27

Query Match 26.5%; Score 3129.5; DB 3; Length 1457;
 Best Local Similarity 46.0%; Pred. No. 3e-288;
 Matches 676; Conservative 207; Mismatches 377; Indels 211; Gaps 28;

```

QY 888 MEERPTHLKLGVSIONLVKVRDGMKVAVDGLALNFEGQITTSPLNGAGKTTMSILT 947
D 1 MEERPTHLPLVVCYVDKLVKYNKDKLANKLSLNLLENQVVSFLGNGAGKTTMSILT 60
QY 948 GLPPTSGATVILGKDISEMSTIRQNLGVCQPHNVLPDMLTVEHMTFYARLGLSEKH 1007
D 61 GLPPTSGATVILGKDISEMSTIRQNLGVCQPHNVLPDMLTVEHMTFYARLGLSEKH 120
QY 1008 VAEEMOALDGLRPSKLSKTSQSGMQRKLSVALAFVGGSKVYLIDEPYAGDPYS 1067
D 121 IKFEDKMIEDLEL-SNKRHSLYQTLSSGMKRKLSVALAFVGGSKVYLIDEPYAGDPYA 179
QY 1068 RRGIVELLKRYQGTIILSTHMDADYLDRIAIIISHGLKCCVSSLFLKNOIGTGY 1127
D 180 RRAIWDILIKYKPGRTIILSTHMDADYLDRIAIIISHGLKCCVSSLFLKNOIGTGY 239
QY 1128 LFLVKKDVSSLSSCRNSSTVSYLKKEDSVSQSSDAGLSDHSDPILTTIDVSAISLI 1187
D 240 LFLVKKDVSSLSSCRNSSTVSYLKKEDSVSQSSDAGLSDHSDPILTTIDVSAISLI 279
QY 1188 RKHVEARLVEDIGHELYTVLPEAKEGAEVLEFHEIDRLSDLGISYISSETLEEI 1247
D 280 RKHVASSLVSTSTSELSTIIPSEAVKKGAFRLFOGLEHSLDHLSSFLGMOTTEEV 339
QY 1248 FLKVAE-----SGVDA-ETISDGLPARRRNRAFQKQSCIRPTE----- 1287
D 340 FLKVAEEDQSLSEADYKESRKDVLPGAEGTLAVGQAGNLARCSLELAQSASLQSSASS 399
QY 1288 -----DDAADPNDSIDIPRESRETLDSGMGKSKSYQVKGKML 1324
D 400 VGSARGEESTGYSDQGYRPLFDMLDDPD--NWSLDENEMALQV--GQSGRKLEGWML 456
QY 1325 TQOQFVALMKRLRLIARRSRKGFQAQIVLPAVFCIALVFSIIPFGKYSLELOPMY 1384
D 457 KMRQHGILLVKRFGHARRNSKALCSQIILLPAFFVCAMTVALSVPEIDDLPLVLSPQY 516
QY 1385 NBOYT-----EVSNDAPF-----DTGTLLELNLATKDPGFGRCM----- 1419
D 517 H-NYQPRGNFIPYANERQERYRLRSPASPOQLVSTFRLPSPGVGATCVLKSPANGSLG 575
QY 1420 -----EGNPI-----PD----- 1426
D 576 PMLNLSGGSRLAARFDSKLESTQGLPLSNFVPPPPSPAPDSVXPVEDSLQAWN 635
QY 1427 --TPQAGEEEMTAP-VQITMDLFONGNMTQMPSPACQSSDKIKMLPVCPPGAG 1483
D 636 MSLPPTAGETWTSAPLRLVHEPVR-----CTCSAGCTGS--CPSVSG 680
QY 1484 LPPPOKONTADILODLGRNISDYLVKTYVQIIAKSLKNKIWNEFRYGGFSLGVSNTQ 1543
D 681 -HPQWRVVTGILDTDIGHNYSYLLFTSDRF-----RLHRYGATTFG--NVQ 726
QY 1544 ALPPSEVDADIKKKKHLKLAKSSADRFNLSLGRFMTGLDTRNNVKKWPNKMWALIS 1603
D 727 KSLPAS-----FGARVPMVAKIARVRAQVLYNNKQYHSNP 763
QY 1604 SFLANINNALRANLQGE-NPSHYGTAFAFNPLNLTQOOLSEVALMTTSVDVLSICV 1662
D 764 TYLNSLNAILRANLPKSKGNPAAYXIITVNHMPMKTSSALS-LDYLLQGTVDVIAIFI 822
QY 1663 FAMSFPASFPVFLDQERKSKAKHLQFTSGVAPVYIWLSENFWDKCNVVPVTLVITIFI 1722
D 1722 FAMSFPASFPVFLDQERKSKAKHLQFTSGVAPVYIWLSENFWDKCNVVPVTLVITIFI 1722

```

```

D 823 VAMSFVPASFVELVAEKSTRAKHLOFVSGCNBEVIYMLANYWMDMLNYLPATCCVILF 882
QY 1723 CFQOKSVSSSTNLPVLLALLLLYGWSTITPLMYPASFVFKIPSTAYVLTSLVILGINS 1782
D 883 VFDPATYSPTNFPAVLSLFLYGNSTITPLMYPASFVFKIPSTAYVLTSLVILGINS 942
QY 1783 VAFVLELFT-DNKLNNINDILKSVLELPPHCLGRLIDMYKNQAMADALEFGE-NRF 1840
D 943 VAFVLELFT-DNKLNNINDILKSVLELPPHCLGRLIDMYKNQAMADALEFGE-NRF 1002
QY 1841 VSPSLMDLVGRNLFMAAVEGVFFLTVLIQYRFFRPPRVAKLSPLNDEDEVRREQ 1900
D 1003 KSPFEMDVLTRGVAMTVEGVFFLTVLIQYRFFRPPRVAKLSPLNDEDEVRREQ 1061
QY 1901 RIIDGGGNDILEFTELKTYRRK--RKPAVDRCVGI--PPCEGGLGVNAGKSTF 1956
D 1062 RVLKGDADNDMVKLENLTKYKSKRIGRIILAVDRDLGVCYBPCEGGLGVNAGKSTF 1121
QY 1957 KMLGDTTVTRGDAFLNKNLSINIEHONMGCPDPAITELLTGRHEVERFALLRGV 2016
D 1122 KMLGDESTTGGEAFVNGHSVLKDLLOQOOSLGCPQDPVDDELTAAREHLQYTRLCI 1181
QY 2017 PEKEVGVGEMALRKLGLVYGEKYNAGNSGKRRKLSMALIGCPPVYFLDEPTTGM 2076
D 1182 PKRDEADVMALEKLELITRYADKPAQTYSGGNKRLSTALIGAPFIDEPTTGM 1241
QY 2077 PKARFLMNCALSVKRGSRVLTSHSMECEBALCTRMALMVGRFCIGSYOHLKNRF 2136
D 1242 PKARFLMNLILDLTKGRSVLTSHSMECEBALCTRMALMVGRFCIGSYOHLKNRF 1301
QY 2137 DGYTIVIRIAGSNPDLKPVDFGLAFPGSVLKEKHRLMLQYLPSSLARIFSLSQ 2196
D 1302 DGYTIVIRIAGSNPDLKPVDFGLAFPGSVLKEKHRLMLQYLPSSLARIFSLSQ 1360
QY 2197 SKRRLHIEDYVSQTTLDYVFNFAKQSD 2227
D 1361 VGVGLGIEDYVSQTTLDYVFNFAKQSD 1391

```

RESULT 5

```

US-08-665-259-25
: Sequence 25, Application US/08665259
: Patent No. 6028173
: GENERAL INFORMATION:
: APPLICANT: Landes, Gregory M.
: APPLICANT: Burns, Timothy C.
: APPLICANT: Connors, Timothy D.
: APPLICANT: Dackowski, William R.
: APPLICANT: Van Raay, Terence J.
: APPLICANT: Klingner, Katherine W.
: TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
: TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME
: NUMBER OF SEQUENCES: 73
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: GENZYME CORPORATION
: STREET: One Mountain Road
: CITY: Framingham
: STATE: Massachusetts
: COUNTRY: United States of America
: ZIP: 01701
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/665,259
: FILING DATE: 17-JUN-1996
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Dugan, Deborah A.
: REGISTRATION NUMBER: 37,315
: REFERENCE/DOCKET NUMBER: IG5-9.1

```

TELECOMMUNICATION INFORMATION:
 TELEPHONE: (508) 872-8400
 TELEFAX: (508) 872-5415
 INFORMATION FOR SEQ ID NO: 25:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1684 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-665-259-25

Query Match 22.4% Score 2642.5; Db 3; Length 1684;
 Best Local Similarity 35.7% Pred. No. 1,2e-241;
 Matches 639; Conservative 300; Mismatches 523; Indels 327; Gaps 48;

534 AGIVETGTTGSGIE-LPHHVKYKIMDIDNERTN-----KIKDGYW----- 574
 111 AAVVEHFENHSKEPLPLAVKHLRF---SYTRRYMWIQTGSPFLKTEEGMHTTSLFPL 167
 575 --DPRADPED---MRYWGFAYLDDVVEQALIRVLTE-----KKTGYWQOMP 624
 168 PNPBPRLTSPDGEPEYIRGFLAVOHAVRAIMEHADAATROLFQRLTVTIKRPY 227
 625 PCYVDIDFLRVSRMPLMTLAWISVAVITIKGIYKEKRLKETMRIMGLDMSILMFS 684
 228 PPIADPPLVAIQYDPLPLLLSTFTYALTARAVOKEKRLLEYMRMGLSMLHWSA 287
 665 WEISLILLYSAGLIVILKLG-----NLPSDPSVVEFSLFAVVTLLQCLISTL 739
 288 WFLLEFLLLLAASFMTLIFCVKVPNAVAVLSRSPSLVFLFLCEALISTSFSPWSTF 347
 740 FSRANILAAAGGIITFLYLPVLCVAMQDYVGFTIKFASLSVPAFGCECEFALEP 799
 348 FSKAMMAAFGLFETTYTPYFVAPRYNMWITLSQKLCCLLSNVAAKMAQILGFEEA 407
 800 QGIGVQWNLFESEPYE-EDGFNLTTSSMLDTELYGVMTVTEAVFPGQYGIIPRYE 858
 408 KGMGIOWMDLL-SPYVNVDDDFCGVGLMLLIDSVLYGLVTWMEAVPGQFVGQWYF 466
 859 PCTKTYMFE-----ESDEKSHPGSNOKRMSLCEMEEPHLKGLYSIOWLVKYYDGM 912
 467 FIMSTYWGCKRAVAKKEEDSDP---EKALRNEYFEPEPDLVAGIKIKLSVVFVGN 523
 913 K--VAVDGLALNFYEQITSPFLGHNAGKTTMSILTGLPPTSGTAVIILCKDIRSEMT 970
 524 KDAVAVRLNLINLEYGQITVLLGHNAGKTTTISMLTGLPPTSGTAVIISYETLSQMYQ 583
 971 IRONIGVQPOHNVLEDMITVEEHIMFYARLKGLSEKHAKAMEDMALDVGIPSSKLKSKT 1030
 584 IRTSLGICPOHDILFDNLFTVAEHLIFYAOLKGLSRKOCPELVKQMLHIGL-EDKMSRS 642
 1031 SOLSGMOKRISVALAFVGGSKVILDEPTAGVDPYSRGTIMELLEYROGRTILSTH 1090
 643 RFLSGMRRKISIGALIASGKVLIDEPITSGMAISRRAIWDLLOKOSRTITVLTHF 702
 1091 MDEADVLCDRILAIISHGKLCVSSSLFKNLQIGTYVLTIVAKKDVESLSSCRNRSSTVS 1150
 703 MDEADVLCDRILAIISHGKLCVSSSLFKNLQIGTYVLTIVAKKDVESLSSCRNRSSTVS 1150
 1151 YLKKEDEVSSOSSDAGLGSOHESPTLITIVSAISNLIRKHVSEKRLVEDIGHELITVLPY 1210
 747 HCNEDP-----ISOLVHHVHPNATLSSAGAEISLFLPR 780
 1211 EAKEGAFVELHEIDRLSDGISYGISSETLEEFILKAAE--ESGVDAETSDGLPA 1268
 781 ESTHR--FEGFLAKLEKKOKKGLIASGASITIMEEVLAVKGLKVDSSMLQIAQ--LPA 836
 1269 ---RRNRARF---DKOSCLRPFTEDDAAPND---SDIPESREITDLSGMGKGSYOV 1319
 837 LQYHERARASDAVDSNLC-----GAMPSPDICALIEEERTAVYKLTNGL----- 881
 1320 KGMKLTQOOFAALLMKRLLIARSRKGFPAQIYPAVFCALVFSLLVPPCKYPSLEL 1379

882 ---ALHCOQFMAMFLKKAAYSWMKMAOVAOVLPTCVTLALL----- 922
 1380 QPMWYNQYTFVSNDAPEDEGTIELLNAITKDPGEFTROMEGNPPIPDTCQAGEEMTTA 1439
 923 -----ATINSELEPDPML--RLTLG-----EYGR 946
 1440 PVEQITMDLFONGMWIMONSPACQCSSDIKMKLPVCPGAGLPPOROKONTADIID 1499
 947 VVPFSVGTSLQDQLESEHLKDALQAG-----QEPREVLGD 983
 1500 LTGRNISDYLKTYVOI IAKSLKNTIWNDFRY---GGFSIGVSTQALPPSQEVNDAI 1555
 984 L-----EEFLI-----FRASVGGGFP----- 1000
 1556 KQMKHKLKAKDSSADRLNSICRFMTGDTGRNNYVWVNMGWHAISSFLVINNALR 1615
 1001 -----ERCL--VAASFVDGERTVYNALFPNOAYSPALAVDNLFLK 1043
 1616 ANIQKENDSHGITAFNHP-----LNLTKQQLSEVALMTTSVDVLVSICVIFAMSPVA 1670
 1044 --LLCG---PHASIVSNFPQPSALQAAKQFNE---GRKGFIALNL--LEAMAPLAS 1093
 1671 SEVFLIOERVSARKHLOFTSGKPIYKLSNFVMDCMYVVPATLVITIFICEQOKSYV 1730
 1094 TETILAVSERAAVAKHVQSGVHVASEWLSALIMDLISFLIPSLLLLVFAKFDVRAF 1153
 1731 SSTNLPLVALLLLGWSITPLPASPAPFVKIPSTAYVLTSVNLFIGINSATFVEL 1790
 1154 RQGMADDTLLLLLYGMAIPLMYLNMFFLGATATYRTLTIFNLISG---AIFLWYT 1209
 1791 ---FTQNLNINIDILKSFELFPHCLRGGLIDWKN-----QAMADALERPG- 1836
 1210 IMRIPAVKVEKSTLDHDFVLVLPNHLGMAVSEYENETRYCTSSVAHYCKYNI 1269
 1837 ---ENRFVSLSWDL--VGRNLFAMAVEGVFELITVLLOYRFFTRPPVNAKL----- 1885
 1270 QYQENY-----AMSARGVRFVASMASCAYLLILFLETNLLORLICALRRRTL 1325
 1886 ---SPLNDEEDYRBRKORILDOGGONDI---LEIKELTKYIRRRKP--AVDRICV 1934
 1326 TELYTRPVLPEODVADRTRILAPSPSLHTPLTIKELSKY--EORVPLADRLSL 1384
 1935 GIPRGCEFGILGVNAGCSTFKMLTGDFTVRCGAFILKNSIILNIEVHOVMGQCOF 1994
 1385 AVQGCDFGLDFGNAGKTTFKMLTGESLTSFGAPFVGNHISDVGVKVRORIGYCOF 1444
 1995 DATFELLTGREHVEFALLRGVPEREKGVEGMAIRKGLVYKGYKAGNTSGGNKRLS 2054
 1445 DALLDMTGREMLVMYARLGRIPERHIGACVENTRLGLLEPHANKLVRTYSGNKRKLS 1504
 2055 TMALLIGSPVVEFDEPTGMDPKARRPLMNCALSVNKEGRSVNLTSHMECEALCTRM 2114
 1505 TGTALIGEPVAVFLDEPSTGMDPVARRLMDTVARARSAGKAITTSHMECEALCTRL 1564
 2115 AIMVNGRFGCLGSVOHLKRGFGDYTIIVRI--AGSNPDLKPVODFGLAPPGSYLKEKH 2172
 1565 AIVWQOQFECCLGSPQHLKSKFGSGYSLAKVQSEDOQALEEFKAFVDTLPFGSVLEDBH 1624
 2173 RNMLOYOLPSSLSLARIPLSLSQSKRLHIEDVSVSOTTLDOVFVNA 2221
 1625 QGVVHYHLPRGRDLSWAKVGLLEKAKKEKGVDDVSVSISLEOVFLSPA 1673

RESULT 6
 US-08-762-500-25

Sequence 25, Application US/08762500

Patent No. 6030806

GENERAL INFORMATION:

APPLICANT: Landes, Gregory M.

APPLICANT: Burn, Timothy C.

APPLICANT: Connors, Timothy D.

APPLICANT: Dackowski, William R.

APPLICANT: Van Raay, Terence J.

APPLICANT: Klinger, Katherine W.
 TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES.
 TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME
 NUMBER OF SEQUENCES: 83
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: GENZYME CORPORATION
 STREET: One Mountain Road
 CITY: Framingham
 STATE: Massachusetts
 COUNTRY: United States of America
 ZIP: 01701
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/762,500
 FILING DATE: 09-DEC-1996
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/665,259
 FILING DATE: 17-JUN-1996
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/US96/10469
 FILING DATE: 17-JUN-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Dugan, Deborah A.
 REGISTRATION NUMBER: 37,315
 REFERENCE/DOCKET NUMBER: 165-9.3
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (508) 872-8400
 TELEFAX: (508) 872-5415
 INFORMATION FOR SEQ ID NO: 25:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1684 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-762-500-25
 Query Match 22.4%; Score 2642.5; DB 3; Length 1684;
 Best Local Similarity 35.7%; Pred. No. 1.2e-241;
 Matches 639; Conservative 300; Mismatches 523; Indels 327; Gaps 48;

QY 913 K--VAVDGLALNEFGQITSLFGLHNGAGKTTTMSLTGLFPPTSTAYILGKDIRSEMST 970
 DB 524 KDRAAVADLNLNLYEGQITSLFGLHNGAGKTTTMSLTGLFPPTSTAYILGKDIRSEMST 583
 QY 971 IRONLGVCPQHNVLFDMLTVEEHIMWFAKLGISEKHVAEMOMALDVGLPSSKLSKT 1030
 DB 584 IRKSLGCPQHDILFDMLTVAEHLFYAQLKGLSRCKPEVKOMLHIGL--EDKWNSSRS 642
 QY 1031 SOLSGMQRKLSVALAVGSSKYIIDEPYAGDPYSRKIMWLLKYGQRTIILSTHH 1090
 DB 643 RFLSGMQRKLSVALAVGSSKYIIDEPYAGDPYSRKIMWLLKYGQRTIILSTHH 702
 QY 1091 MDEADVLGDRITAIISHGKLCVSSSLFKLNGQGYTLTVKCKDVSSLSGSSSTVS 1150
 DB 703 MDEADVLGDRITAIISHGKLCVSSSLFKLNGQGYTLTVKCKDVSSLSGSSSTVS 746
 QY 1151 YLKREDSVSSSSDAGLGSDESPTLTIDVSAISNLIRKHVSARLVEDIGHELTVLPY 1210
 DB 747 HCNPE--ISQVHNHVPNATLESSAGALSFLPR 780
 QY 1211 EAKEGAFVELFHEIDRLSDLGISYGISPTLEELFLKVAE--ESGVAEHSDGTLPA 1268
 DB 781 ESTHR--FEGLPALKERKQKELGASFGASITTWEEVFLRVGLVDSMDIQAIQ--LPA 836
 QY 1269 --RRNRRAFG--DKOSCLRPFTEDDAADPN--SDIPRESRETLDSGMDGKSGYOV 1319
 DB 837 LQYHERRASDMAYDSNLC--GAMPDSGIGLIEERTAYKLNGL-- 881
 QY 1320 KGMKLTQOOEVALLMKRLLIARRSRKGFPAOIVLPAVFCIALVSLIVPFCKYPSLEL 1379
 DB 882 --ALHCQEFWAMFLKKAASWREMKVAQAQVLPVLCVTLAL-- 922
 QY 1380 QPMWYNEDYTFVSNDAPEDGTLELLNALTKPDGFCRGKMGNDPDTQCAQEEEMTFA 1439
 DB 923 --ATWSELEFDDPML--RLTLG--EYGR 946
 QY 1440 PVPTQINDLFQNGMWTQNPSPACOCSSDKIKKMLPVCPGAGGLPPQKONTADILAD 1499
 DB 947 VVPSVPSGTSGQLQQLSEHLKDALQAG--QEPREVLGD 983
 QY 1500 LTGRNISDYLYKTYVOIILASLKKIWNERY--GGSFLGVSNTQALPDSQEVNDAL 1555
 DB 984 L--EEFL--FRASVEGGGPN-- 1000
 QY 1556 KQMKHLKLAOSSADRPILNSLGRFMTGLDTRNNYKWPFNKKHATISFLVINYNAILR 1615
 DB 1001 --ERCL--VAASPRDVGERTVYNALEPNQAIHSPATLAVAVDNLLEK 1043
 QY 1616 ANLQGENSHYGITAFNHP--LNLTKQOLSEVALMTTSVDVLVSICVIFRASFVPA 1670
 DB 1044 --LLCG--PHASIVSNFQPRASALQAAKQFNE--GRKGFIDALNL--LFAFATLAS 1093
 QY 1671 SFVVFLEIDERYSKAKHLOFISGVKPVYIWMLSNFVDMCNVYVATLVITIFICPOOKSYV 1730
 DB 1094 TFSILAVASERAVQAKHVOFSGVHVASFVMSALMDLISFLIPSLILLVYKFAEDVAF 1153
 QY 1731 SSTNPLVALLLLLYGMSITPLMPASPFVKIPSTAAVVLVSVLFTIGISVATFVLEL 1790
 DB 1154 RDGHMADTLILLLXGNAIIPMLTLMNFPLIGATATATRLTIFILISGT--ATFLMWT 1209
 QY 1791 --FTDNKLNINIDILKSVFLIPFPCIGRGLIDVKN--QAMADALBERG-- 1836
 DB 1210 IMRIPAVKLELSKTLIDHVLVLPNHGCMGVSSFEVETRYRSTCSSEVAABHKKYNI 1269
 QY 1837 --ENKREYSPISWDL--VGRMLFAMAVEGVFFLITVLIQYRFRIRPPVNAKL-- 1885
 DB 1270 QYQENFY--AMSAFGVGRFVASMAGCAVYLLLEFLIETNLIQRLGILCALARRRTL 1325
 QY 1886 --SPUNDEDEVYRERORITLDGGGONDI--LEIKELTKYRKKRP--AVDRICV 1934
 DB 1326 TELTYRMVLPEDODVADERIRILAPSPDSLHTPLIKELSKYV--QORVPDLAVDSL 1384
 QY 1935 GIPGECFGLLVNGAGKSSTFKMLTGDITVTRGDAPFLNKNSTLSINLIEVHONKGYCPOF 1994

```

Db      1385 AVOKGGCFLLEFNGAGKTTTFRKMLTGESLISGDAFVGAGHISSDYKVRORICYPQF 1444
QY      1995 DATTELLTREHVEFFALLRGVEKEVGVGEMAIRKILGLVYKRGVANGNSGCKRRLS 2054
Db      1445 DALLDHMTGRENLMVYARLRIGTPEHNIGACVNTLRGLLEPHANKLVYRTSGCKRRLS 1504
QY      2055 TMAALIGRPVYVLEDEPTTGMDPKARFRLNWCALSVEKGRSVLTLSHMECEALCTRM 2114
Db      1505 TGIALIGEPVAVIFLEDEPSTGMDPVARRLMDTVARARESGKAITTSHSMECEALCTRL 1564
QY      2115 AAMVNGREFCIGSYOHLKRRFGDGYTIYVRI--AGSNDLKRVQDFGLGAPGSVLKREKH 2172
Db      1565 AAMVNGREFCIGSYOHLKRRFGDGYTIYVRI--AGSNDLKRVQDFGLGAPGSVLKREKH 1624
QY      2173 RMLQYQPLSSLSLARIFSLISOSKRLHEDYSVSTQTLDOYVNEFA 2221
Db      1625 QCMVHYHLPGRDLISMAKVGILKAKERYGVVDYSVSQISLEQVFLSFA 1673

```

RESULT 7

```

; Sequence 75, Application US/08762500
; Patent No. 6030806
; GENERAL INFORMATION:
; APPLICANT: Landes, Gregory M.
; APPLICANT: Burn, Timothy C.
; APPLICANT: Connors, Timothy D.
; APPLICANT: Dackowski, William R.
; APPLICANT: Van Raay, Terence J.
; APPLICANT: Klinger, Katherine W.
; TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES.
; TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME.
; NUMBER OF SEQUENCES: 83
; CORRESPONDENCE ADDRESS:
; ADDRESS: GENZYME CORPORATION
; STREET: One Mountain Road
; CITY: Framingham
; STATE: Massachusetts
; COUNTRY: United States of America
; ZIP: 01701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/762,500
; FILING DATE: 09-DEC-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/665,259
; FILING DATE: 17-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/10469
; FILING DATE: 17-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Dugan, Deborah A.
; REGISTRATION NUMBER: 37,315
; REFERENCE/DOCKET NUMBER: IGS-9-3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508) 872-5415
; TELEFAX: (508) 872-8400
; INFORMATION FOR SEQ ID NO: 75:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1704 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-762-500-75

```

```

Query Match      22.4%; Score 2642.5; DB 3; Length 1704;
Best Local Similarity 35.7%; Pred. No. 1,3e-241;

```

```

Matches 639; Conservative 300; Mismatches 523; Indels 327; Gaps 48;
QY      534 AGIYETITPGSIE-LPHNKKYKIRMDIDVETRN-----KIDGVW----- 574
Db      131 AAVFEHPFHNSKEPLPLAKVHLR--SYTRRNMTQTGSEFLKETEGMHTTSLFPL 187
QY      575 --DPGRADPEED--MRVWGGFAVLQDVEEQAIRVLTGTE-----KKTGYWQOMP 624
Db      188 FPNPGPRELISPGGEGRYIREGFLAVQHAVDAIMEYHADATROLPLRLVYTIKRPFY 247
QY      625 PCYVDDIFLVRMSKSMPLFMTLAMITYSAVVIKGIYKEEARLKETMRIMGDNSTLWFS 684
Db      248 PPLADPFLVAIQYPLPLLLLSFTYALTARAVQOEKERRLEKEXMRMGSLSWLHMSA 307
QY      685 WFLISLPLVLSAGLVVILKLG-----NLTPSDSVVVFVFAVAVTIIOCFILSTL 739
Db      308 WFLFLFLFLILIASFMILLFCVKKPRNVALSDDPSVLATFLICFAITISFSFVSTF 367
QY      740 FSRANLAACGGIILYFLYLVLCVAMQDYGETIKIFASLSPYAFGGEYFALFEE 799
Db      368 FSKANMAAAGGFLYFFTYIPYFFVAPRYNMWTLISQKLCISLNVAMAMGALIGKFEA 427
QY      800 QGTGVQMDNLFESPYE-EDGFNLTTSIMMLPDTFLYGMVTVIYAVFGQYGRPPTF 858
Db      428 KGMGLQMRDL--SPVAVDDPCFGQVLGMLLDSVLYGLVTVYMEAVFPFGQVDPWYF 486
QY      859 PCIKSYWFE-----ESDEKSHSGSNOKRMSRFICMEEPFHILKGVSTONLVKVRDM 912
Db      487 FIMPSYWCGRPAVAKKEEDSDP---EALNREYFEAPREDLVAGIKIKLSKPYRGN 543
QY      913 K--VAVDGLANFYEQITSFIGNAGAKTTWSTILTG-LPPTSGTYILGDRISENST 970
Db      544 KDRAAVRDLNLTNLEQGITVLLGNHAGAKTTLSMLTG-LPPTSGRAYISGEYSQDMVQ 603
QY      971 IRONLVCQPHNVLFPMLEVEHIMFYARLKGLSEKHNVAEEMOALDVLGPPSLKSKT 1030
Db      604 IRKSLGICQPHDLFNLVLAHELYFAOLKGISRKCCEEVKQMIHLIIG-EDRWNS 662
QY      1031 SOLSGMORKLSVALAFVGSKVYIIDEPTAGVDVPSRRGIWELLKYROGRTIILSTHH 1090
Db      663 RFLSGMRRKRLSIGILLIAGSKVILIDEPTSGMDAISRAIMDLLQQRKSDRTIYLTTHF 722
QY      1091 MDEADVIGRIATISHGKLCYCGSSIFLKNQIGTYIYTLVKKDYESSLSGCRNSSYVS 1150
Db      723 MDEADLLGRIALIMAGELQCCGSSLEFLKQYKAGYHMTLVKE-----P 766
QY      1151 YLKKEDSVSQSSSDAGLGSDESDTLTIDVSAISNLIRKHYSEARLVEDIHELTYVLY 1210
Db      767 HCNPED-----ISLVHHHVNATLLESSAGNELSFILPR 800
QY      1211 EAKBGAVELPHEIDRLSDGISYIGSETTLEIFLVAE--ESGVDAETSDGLTLP 1268
Db      801 ESTHR--FEGLEPAKLEKKQKELGIASFASITTEMEVFLRVKLVDSMDIOAIQ--LPA 856
QY      1269 ---RRNRAPFG--DKQSLRPFTEDDADPNV---SDIDPSRETLDSMGDGSQYV 1319
Db      857 LQYQERRASDMAVDSNLC-----GANDPSDYGALIEERTYAVKNTGL----- 901
QY      1320 KGMKLTQOQFVALMLKRLDIARRSRKGFPAQIVLPAVVCIALVSLIVPFGKYSLEL 1379
Db      902 ---ALHCQGFWMFLKKAAYSREKKMVAQVLYLVCTYALL----- 942
QY      1380 QRMVYNEQYTFVSNAPDETGTLELLMALTKDPGGGTROMBSNPJPDTPCAGEEETPA 1439
Db      943 -----AINVSSLEFDDPM--RLTLG-----EYGR 966
QY      1440 PVQTIMDLFQNGNMTWQNPSPACOCSSDKIKMLPVCPGAGGLPPQRRKONTADILQD 1499
Db      967 VVFPSPGTSQGLQSLSEHLKDALQAE-----QDEPREVLGD 1003
QY      1500 LTGRNISDYLVKTYVOIATKSLKNKIWNPEFY-----GAFSLGVSNTOALPPSQEVDN 1555
Db      1004 L-----EFLI-----FRASVEGGGFN----- 1020

```



```

QY 1556 KKKKKHKLAKDSADREFNSLGRFMTGLDPNNKVMFNKMKHATSSPLINNALR 1615
DB 1021 -----ERCL--VAASFROGERTYVNALENNQAHSPALAAVDNLLFK 1063
QY 1616 ANLOGENSHYITAFNHP-----LNLTKOQSEVALMTTSVDVLSICVIFAMSEVPA 1670
DB 1064 --LLCG---PHASIVVSNPPQSRSLQAQKQFNE---GRGFDIALNL--LFMAAFIAS 1113
QY 1671 SFVVELIOBRKSKAHLOFISGVKVIITWLSNFWDMCNVYVPAVLITITFCQOKSYV 1730
DB 1114 TFSIIAVSERAAQKHOVQSVGVHVASFMSALMDLISFLPSLLLVFKAPDVRAFT 1173
QY 1731 SSTNPLVALLLLTGMSTPLMPASPFVKIPSTAYAVLTSVNLFIGINSVAFVEL 1790
DB 1174 RCGHMAOTLLLLLTGWAITPLMTLMNEFFLGATATATYTRLTITFNLSGI---ATFLVVT 1229
QY 1791 ---FTDNKLNINDILSKVFLIFPHCLGRGLIDVKN-----QAMADALERFG- 1836
DB 1230 IMRIPAVKLEELSKTLDHVELVLPNHCIGMAVSSFEYENETFRYCTSSSEVAHCKKYN 1289
QY 1837 ---ENRFVSPLSMDL--VGRNLFMAVEGVVFLITVLIOYFFLRPPRVAKL----- 1885
DB 1290 QYQENFY---AMSAPGVGRFVAMASCAAYLLFLIETNLLORGLICALLRRRTL 1345
QY 1886 -----SPINDEDEDVRRRORILDGGGNDI---LEIKELKIVRRRRKP--AVDRICV 1934
DB 1346 TELYTRMPLPDPDOYADERTILAPSPSLHTPLLIKELSKV--EORVPLAVDRISTL 1404
QY 1935 GIAPGECFGLGVNGAGKSTFKMLGDTTVTRGDAFLNKSILSNIEHVOMNMGCPQF 1994
DB 1405 AVQKGEFCGLGFNGAGKSTFKMLGDTTVTRGDAFLNKSILSNIEHVOMNMGCPQF 1464
QY 1995 DATETLLTGREHVEFPALLRGVPEKEVGKVGEMAIRKIGLVYKGRKAGNYSCKKRLS 2054
DB 1465 DALDMDTREMILVMYARLGRIGPERHIGACVENTLRGLLEPHANKLVRTYSCKKRLS 1524
QY 2055 TMAALIGBPVVELDEPTGMPKARFLMNCALSVYKGRSVLVLSHMECEALCTRM 2114
DB 1525 TGAALIGBPVVELDEPTGMPKARFLMNCALSVYKGRSVLVLSHMECEALCTRM 1584
QY 2115 AIMVNGRFPCLSGVOHLKNRFGDGYTIVRT--AGSNPLKPVODFGLAPGSLKEKH 2172
DB 1585 AIMVNGRFPCLSGVOHLKNRFGDGYTIVRT--AGSNPLKPVODFGLAPGSLKEKH 1644
QY 2173 RMLVQOLPSSLSLARIFSIISSQSKRLHIEDYSQOTLLDOVYNFA 2221
DB 1645 QGMVHYHLGRDLSMAKVFGLLEKAKKGYVDVYSQISLSEOVLFSA 1693

RESULT 8
US-08-583-276-19
: Sequence 19, Application US/08583276
: Patent No. 5837536
: GENERAL INFORMATION:
: APPLICANT: McDonagh, Kevin T.
: APPLICANT: Nienhuis, Arthur
: APPLICANT: Tolstoshev, Paul
: TITLE OF INVENTION: IMPROVED EXPRESSION OF HUMAN
: TITLE OF INVENTION: MULTIDRUG RESISTANCE GENES AND IMPROVED
: NUMBER OF SEQUENCES: 19
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Carella, Byrne, Bain, Gilfillan,
: STREET: 6 Becker Farm Road
: CITY: Roseland
: STATE: New Jersey
: COUNTRY: USA
: ZIP: 07068
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5 inch diskette
: COMPUTER: IBM PS/2

```

```

: OPERATING SYSTEM: PC-DOS
: SOFTWARE: DNA V2
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/583,276
: FILING DATE: 05-JAN-1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/332,444
: FILING DATE: 31-OCT-1994
: APPLICATION NUMBER: 07/887,712
: FILING DATE: 22-JAN-1992
: INFORMATION FOR SEQ ID NO: 19:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1280 amino acids
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: Linear
: MOLECULE TYPE:
: DESCRIPTION: protein
: US-08-583-276-19

```

```

Query Match 2.9%; Score 340.5; DB 2; Length 1280;
Best Local Similarity 18.7%; Pred. No. 4.1e-22;
Matches 272; Conservative 201; Mismatches 470; Indels 509; Gaps 56;

```

```

QY 776 KIFASLSPVAFGCGCYFALFEEOGIGVOMDLFESPEDEDFNLTTSSMMLPDTFL- 834
DB 256 EVLAIRIVAFGQKKELEKYNK-----NLEEA--KRIGKKAITANISIGAFL 305
QY 835 ---YGVATWY-IEAVPEGOYIPR---PWYFCTKSYMFGES----- 870
DB 306 IYASVALAFVGTTLVLSGEYSIGQVLTFFSVLIGAFSVGOASPTEAFANRGAAYEI 365
QY 871 ----DEKSHPSNCKRNSICEEPEPHLKIGVSIQWLVAVYVDGMKVAA--DGLANFYE 925
DB 366 FKIIDNPKSIDYSK-----SGHKPNKIGLNEFRVNHSTYSRKEVKILGLNLKVS 419
QY 926 GQITSFHNAGCTTMSITGLFPEPTGTAYILGDIRS--EMSTIRQMLGVCPOHNL 984
DB 420 GQYVALVNSCGSGSTVOLMORLYDPTBEGVSVYDGDITINRFLRELIGVSGPVL 479
QY 985 FDMLTVEEHWIYFARLKLSEKHYKAEQMAID--VGLP---SSKLSKTSQLSGGMOR 1039
DB 480 F-ATTIENI--RYGRONVYMDIEKAKVAMAYDFIMKLPKEDTLVGERGAOLSGQOK 537
QY 1040 KLSVALAVYGSKVYIIDEPTAGVDPYSRGIWELLKRYOGRTIISTHMD---EADV 1096
DB 538 KIALARALVNRPKTILLDEATSLDTESEAVVOVALDKARKGRTTYIAHRLSTVRNADY 597
QY 1097 LG--DRIAIISHGKLCVGSLSFLKNOLGTGYTLTVKKDV---ESSLSGRSSSTVSY 1151
DB 598 IAGFDGCVIVKSG-----NHDELKMKK--GYFKLVMTQAGNEVLEMAADESKS--- 646
QY 1152 LKEDSVSOSSDAGLSDHESDITLIDVSAISLIRKHNSEARLVEDIGHETLYVLYPE 1211
DB 647 --EIDALEMSSNDR-----SSLRK----- 665
QY 1212 AAKBGAFFELHEIDRLSDGISSTGSETTLEIFLKAABESGVDAERTSDGLPARRN 1271
DB 666 -----RST 668
QY 1272 RRAFDKQSLRPPTEDDAADPNDSDIDPESRETDLLSGMDKSGYOVKMKLQOQFVA 1331
DB 669 RRSVAGSQAOQRKLSTKCAL--DESIPP-----VSFWRIKMLWTE 707
QY 1332 LMKRLLIARRSKRGFPQOYIYLPVAVFCIALVESLIYPPCKTYSLELDQWMTWEOYTFV 1391
DB 708 --WYFVY-----GVFCALINGGIQAPAFATIFSKII----- 736
QY 1392 SNDAPEDVTGLELLNALIKDKPGFGRMGEGNPIPDTPQAGEEMTTAPVPOTIMDLFON 1451
DB 737 -----GVFTF----- 741

```

```

QY 1452 GNMWQNPSPACOCSSDKIKKMLPVCPRGAGLPPRQKONTADILQDITGNISDIYLVK 1511
Db 742 -----IDDETRKRONSNLFSL-----PLAL 762
QY 1512 TYVQIIAKSLKKNKWNEFRYGGFSLGVSNTOALPPSOEVNATIQMKHKLKLAQSS-A 1570
Db 763 GIISFT-----FLOGFTFGKA-----GELLTKRLKRYMFRSLRQDVSMF 804
QY 1571 DRELNSLGRPMGTDRNNVVKWFNNKGMHAISFLVNNALRANLQKGNPSHYGT 1630
Db 805 DDPKNTGALTTRL-ANDAAY-----KG--AIGSLAVITQNI--ANLGTG-----II 849
QY 1631 AENHPLNLTKQOLSEVALMTTSVDLVISCVTFAMSEVPASVAVVLIDERSKAKHIOFI 1690
Db 850 SFTYGWLT-----LULLAIVPIALAGVEMK-----MLSQALDKKKELE-- 891
QY 1691 SGKPVYIWLNSFVDMCMNVVVPATLVIIIFLCPOK---SVYSTNIPVALL--LLLY 1745
Db 892 GAGKIATEIENF-----RTVVSILTOQOKFEHMAQSLQVPRNSLEKKAHIF 938
QY 1746 G--WSTTPMTYASFEVKIPSTAYVVLTSVNLFIGINGSVATFVLEFTDNKLNINDIL 1803
Db 939 GTFSTQAMMYFSYAGCFRFGAYLVA-----HKLMSFEDVL 975
QY 1804 KSVFLIFPHFCRGLIDKVNQAMADALERPGENRFV-----SPLSMDIVGRNLFAMAY 1858
Db 976 -----LVFSAVVEGAMAVGVSSFPADYAKAKISAHIIIMILEKTPILDSYSTEGLMPTL 1031
QY 1859 EGVVEFLITVLIOYREFTRP-RVNAKLSPLNDEDEDVRRERQRIIDGGONDILEIKEL 1917
Db 1032 EGNVTEGEVY--FMYPRPDIPVLOGLS-----LEVKK- 1062
QY 1918 KTIYRRKKRPAYDRICVGPPECEGGLGVNAGKSTFEKMLTGTTVTRQDAPLANKSI 1977
Db 1063 -----GOTLALVSSSGKSTIVQLLERFYDPLAKVLLDCKEI 1101
QY 1978 LS-NIHEVQNMGYCPO---FD-----ATTELLTGREHEFFALLRGV 2016
Db 1102 KRLNOMLRAHLGIYSOEPIIFDCSIAENIAGDMSRVVSOEIVRAKENIHAFFIESL 1161
QY 2017 PEKEVGKGEKMAIRKLGIVKGEKTAAGNYSGGNKRLKSTAMALIGCPVAVLDEPTGMD 2076
Db 1162 PKRYSTKVGDKGTQ-----LSGQOKORIAARALVQPHILLDEATSMALD 1207
QY 2077 PARRELMNCALSVYKEGRSVVLTSHEMECECALCTRMAIYVNGFRCLGVOHLKKNFG 2136
Db 1208 TESEKVVGE-ALDKAREGRTCIIVAHRLSTION-ADLIVFQNGKVKHEGTHOULLAQKG 1265
QY 2137 DGYTIVRIAGS 2148
Db 1266 IYFSMVSVQAGT 1277

```

RESULT 9

```

5206352-4
: Patent No. 5206352
: APPLICANT: Roninson, Igor B.; Pastan Ira H.; Gottesman,
: Michael M.
: TITLE OF INVENTION: COMPOSITIONS FOR CLONES CONTAINING DNA
: SEQUENCES ASSOCIATED WITH MULTIDRUG RESISTANCE IN HUMAN CELLS
: NUMBER OF SEQUENCES: 4
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/622,836
: FILING DATE: 24-SEP-1990
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 892,575
: FILING DATE: 01-AUG-1986
: APPLICATION NUMBER: 845,610
: FILING DATE: 28-MAR-1986
: SEQ ID NO: 4:
: LENGTH: 1280
5206352-4

```

```

Query Match 2.9%; Score 339.5; DB 6; Length 1280;
Best Local Similarity 18.7%; Pred. No. 5, 1e-22;
Matches 272; Conservative 201; Mismatches 470; Indels 509; Gaps 56;
QY 776 KIFASLLSPVAFGGCGCYFALFEEOGIGVQWMDLFPSPVEEDGFNLITTSIMLPTFL- 834
Db 256 EVLAIKRTVIAFGQKKELERYNK-----NLEEA--KRIGIKAIYANISIGAAFL 305
QY 835 ----YGYMTWY-IEAVPQGYGIPR--PWYFPCTSYMFGEES----- 870
Db 306 IYASVALAFVNGTTLVLSGEYSIGQVLTFVFSVLIGAFSGQASPSITEAFANRGAAYEI 365
QY 871 ----DEKSHPGSNOKRMSSEICMEBEPTHLKGVSIONLVKVVYKDGKAVV-DGLALNFEY 925
Db 366 FKIIDNKPISIDYSK-----SGHKPDNIKGNLEFRNVHSPSRKEVRLKGLNKLKQVS 419
QY 926 GQITSPFLHGAGKATTTMSILTGLEPPTSTAYVILKQDIRS-EMSTIRONLIGCPQHNVL 984
Db 420 GQYVALVGNSSCGKSTTVQLMORLIDPTEGMSVSDQDRTIIVRFLREIIGVSOEPLV 479
QY 985 FDMLTVEHITWYFARLKGISEKHVKAEMQALD--VGLP--SSKLKSTQSLSGGMOR 1039
Db 480 F-ATTIAENI-RYGRENVITMDEIEKAVKRNAAVDYIMKLPKHPDVLVGRGAQLSGQGQ 537
QY 1040 KLSVALAFVGGKSVVILDEPTAGVDPRYKRGITWELLKTRQGTTLTSHHND--EADY 1096
Db 538 RIARLARLVNRPKTLILDEATISALDTESEAVVOVALDKARKGRTTIVIAHRLSTYRNADV 597
QY 1097 LG--DRIATISHKCLCVCSSLEFLKNOLGTGYTLTVKRDV--ESSLSGRRSSSTVSY 1151
Db 598 IAGDDQVITYEKG-----NDELMAKER--GIYFKLVYQTAGNEVELMAADESKS--- 646
QY 1152 LKREDSVQSSDAGLSGDSHESDPTLTIDVSAISNLRKHVSEARLVEDIGHELTVLYPE 1211
Db 647 --EIDALEMSSNSDR-----SSLIRK----- 665
QY 1212 AAKEGAVELFHEIDRLSDLGSSVSGISBTLEELFKVAEESGVDAETSDGTLPARN 1271
Db 666 -----RST 668
QY 1272 RRAFGRQSCIRPTEDDADPNDSIDIPESRETDLLSGMDKGSYQVGMKLTQOQFVA 1331
Db 669 RRSVRSQAQDRKLSTPKEL--DESIPP-----VSFWMIMLNLTLE 707
QY 1332 LMKRILLIARSKKGFPAQIVLPAVVCIALVPSLIVPPFGKYPISLEDPQMYTNEQYFV 1391
Db 708 --WPYFVY-----GVFCALINGILOPARAIIIFSKLI----- 736
QY 1392 SNDARPDGTLELLNALTKRDPGFGTRCMENPIPDTPQAGEEMTTAPVQTIMDLPON 1451
Db 737 -----GVFTT----- 741
QY 1452 GNMWQNPSPACOCSSDKIKKMLPVCPRGAGLPPRQKONTADILQDITGNISDIYLVK 1511
Db 742 -----IDDETRKRONSNLFSL-----PLAL 762
QY 1512 TYVQIIAKSLKKNKWNEFRYGGFSLGVSNTOALPPSOEVNATIQMKHKLKLAQSS-A 1570
Db 763 GIISFT-----FLOGFTFGKA-----GELLTKRLKRYMFRSLRQDVSMF 804
QY 1571 DRELNSLGRPMGTDRNNVVKWFNNKGMHAISFLVNNALRANLQKGNPSHYGT 1630
Db 805 DDPKNTGALTTRL-ANDAAY-----KG--AIGSLAVITQNI--ANLGTG-----II 849
QY 1631 AENHPLNLTKQOLSEVALMTTSVDLVISCVTFAMSEVPASVAVVLIDERSKAKHIOFI 1690
Db 850 SFTYGWLT-----LULLAIVPIALAGVEMK-----MLSQALDKKKELE-- 891
QY 1691 SGKPVYIWLNSFVDMCMNVVVPATLVIIIFLCPOK---SVYSTNIPVALL--LLLY 1745
Db 892 GAGKIATEIENF-----RTVVSILTOQOKFEHMAQSLQVPRNSLEKKAHIF 938
QY 1746 G--WSTTPMTYASFEVKIPSTAYVVLTSVNLFIGINGSVATFVLEFTDNKLNINDIL 1803

```

```

Db 939 GTFSTQMMWFYSVAGCFRCFAYLVA-----HKIMSFEDVL 975
QY 1804 KSVFLIFPHFCGRGLIDMKVKNQAMADALERTGENRFV-----SPLSMDLYGRNLFANAV 1858
Db 976 ---LVFSAVVEGAMAVGVSSSFADPYAKAKISAAHIIIEKTPLLIDYSTEGIMPRTL 1031
QY 1859 EGVFFLLVLVLYQFFFIIPR-PVNAKISPLNDEDEDERERORILLDGGGNDIEIKEL 1917
Db 1032 EGVNVEGEV---FVNPFRPDIPVLOGLS-----LEVKK- 1062
QY 1918 TKYRRKRPAVDRIQVGIPEGEFCLLGVNGAGSSFFKMLTGTPTVTRGDAPLKNKSI 1977
Db 1063 -----GQTLAVGSSGCKSTVQLLEFYDPLAGKVLDDGKEI 1101
QY 1978 LS-NIHEVQNMNGYCPQ---FD-----AITELLTGHEHVEFFALLKGV 2016
Db 1102 KRLNVOMLFAHLIGVISOEPILFDCSIAENIAYGDSRVVSOEIVRAKKEANITIAFIESL 1161
QY 2017 PEREYKVGEMAIRKLGIVKYGEKYAGNYSGNKRRKSLTAMALLIGPPVFLDEPTTGM 2076
Db 1162 PKYSTRKVDKGTQ-----LSGQKQRIARLALVQPHILLDEATSLD 1207
QY 2077 PARPFLMNCALSVYKEGHSVLTSHSMECEALCTRAIMVNGRPFCLGYSYOHUKNRF 2136
Db 1208 TESEKVVQE-ALDKAREGRTCIIVIAHRLSTION-ADLIVFQNGHVKEHGTQOLLAKG 1265
QY 2137 DGTIVVRIAGS 2148
Db 1266 IYFSMVSVOAGT 1277

```

RESULT 10

US-08-784-649A-2
Sequence 2, Application US/08784649A
Patent No. 5830697

GENERAL INFORMATION:

APPLICANT: SAKIC, Branislav I
APPLICANT: Chen, Gang
TITLE OF INVENTION: P-GLYCOPROTEIN MUTANT RESISTANT TO
TITLE OF INVENTION: CYCLOSPORIN MODULATION
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 2200 Sand Hill Road
CITY: Menlo Park
STATE: CA
COUNTRY: USA
ZIP: 94025

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/784,649A
FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Sherwood, Pamela J
REGISTRATION NUMBER: Reg. No. 5830697 36,677
REFERENCE/DOCKET NUMBER: 06037/007001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-322-5070
TELEFAX: 415-854-0875

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 1279 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-784-649A-2

Query Match 2.9%; Score 337; DB 2; Length 1279;
Best Local Similarity 18.7%; Pred. No. 8,7e-22;
Matches 272; Conservative 200; Mismatches 471; Indels 508; Gaps 56;

```

QY 776 KIFASLLSPAFCGCGCYFALFEBQIGVQVQDNLFESVPEEDGFNLTTISMLPFTFL- 834
Db 256 EVLAIRITVIAFGQKKELERYNK-----NLEEA--KRIGIKRAITANISIGAFL 305
QY 835 ---YGVMTWY-IEAVFPQGYIPR--PWYFPCTKSWPEES----- 870
Db 306 IYASYALAFWGTTLVLSGEYSIGQVLTVEFVLGASVGOASPSLEAFANGAAYEIP 365
QY 871 ---DEKSHPGSNOKRMSEICMEEPETHLKLGVSTIONLVKYRQGMKAVV-DGLALNFEY 926
Db 366 KIIDNKSIDSYSK-----SGHKPDNIKLNLEFRNHFYPSRKKEKILKGLNLVQSG 419
QY 927 QITFELHNGAGKTTTMSILTLGFPPTSGTAYILCKDIRS-EMSTYRONLGVCPQHNVLF 985
Db 420 QTVLVGNSGCGKSTVQLMQRLYDPREGWVSDGQDITINVRFLREILIGVSOEPIVF 479
QY 986 DMLTVEEHIMFYARLKLGLSEKHVKAEMQALD--VGLP---SSKLKSTQSLSGGMOK 1040
Db 480 -ATTIAENI-RYGRENVYTMDEIEKAVKEANAYDITMKLPKFDLVGEKAQLSGGOKR 537
QY 1041 LSVLAELVGGSKVYILDEPTAGVDPYSRGIWELLKYRGRTILSTHMD--EADVL 1097
Db 538 IAIARLALVRNPKIILLDEATISALDTESEAVYQVALDKARGRTIYIAHRLSTVRADVI 597
QY 1098 G-DRIATISHGKLCVYSSSLFKNQLOGTYLTVKKVD---ESSLSGRNSSTVSYL 1152
Db 598 AGFDDGVIVKEG---NHDELMEK--GIYFKLVMTQAGNEVELDENADESKS---- 645
QY 1153 KKEDSVSSQSSDAGLGSDHESDITLIDVSAISNLIRKHVEARLVEDIGHETLYVLYEY 1212
Db 646 -EIDALEMSSNDSR-----SSLIIR----- 664
QY 1213 AKEGAFVELFHEIDRLSLDGLISSYGISSETLLEIFLKVAAESGVDAETSDGTLPARNR 1272
Db 665 -----RSTR 668
QY 1273 RAFGDKQSCIRPFTEDDAADPNDSIDIPESRETDLSGMDGKGYOVKMKLQOQFVAL 1332
Db 669 RSVGSOAQODRKLSTKEAL--DESIPP-----VSFWRIMKLNLTE- 706
QY 1333 LMKRLLIARRSRKGPFAQIVLPAFVCIALVESLIVPFGKYPSELQPMVYNQYTFVS 1392
Db 707 -WPIYFVV-----GVFCATINGGLQAPAFALIFSKIT----- 735
QY 1393 NDAPEDTGLELNLALTKDPGFGTRCMEGNIPDTPCOAGEEEMTAPVQTIIDLFPQNG 1452
Db 736 -----GVFTR----- 740
QY 1453 NWTMONPSPACQSSDKIKKMLPVCPRGAGLPPRQKONTADILQDLTGRRNISDYLVKT 1512
Db 741 -----IDDPETKRQNSNLFSL-----FALG 762
QY 1513 YVQIIAKSLKNNKIMWNEFRYSGFSLGVSNIOALPPSQEVNDAIKOMKHLKAKDS-AD 1571
Db 763 ITSFIT-----FFLOGFTFGKA-----GELTLKRLKYMWFRSMKRDVSWFD 804
QY 1572 RFLNSLGRFMTGLDTRNNKQVWNNKGWHAISFLVYINNAILRANLOKGENPSHYGITA 1631
Db 805 DPKNTTGALTTRL-ANDAAQV---KG-ALGSRILAVITQNI--ANGTGL-----IIS 849
QY 1632 FNHPLNLTQKQOLEVALMTTSDVNLVSLCVTFAMSEVPASVYVLIQERKSKAHLQFIS 1691
Db 850 FLYGQMLT-----LALLAIVPIIALAGVENK-----MLSQALDKKKELE--G 891
QY 1692 GVKPVIYWLNSFWMQMCVYVVPATLVIIIFICFOOK---SVYSTNLPVYALL--LLYG 1746
Db 892 AGKIATEIENF-----RTYVSLTORCKFEMHYAQSLQVYRNSLKAHIFG 938

```



```

Db      892 -GAGGIAEAIENF-----RTVVSILQEQKEFHMVYQSLQVYRNSLRKAI 937
QY      1745 VG--WSTIPLMVPASFEVKIPSTAVVYVLTSLNLTGINGSVATVLELTNTKLNINDI 1802
Db      938 FGITTSFTQAMMYFYIACCFREGATLVA-----HKLMSEEDV 974
QY      1803 LKSVLEPFPHEFLGRGLIDMYKNQAMADALERFGENRFV-----SPLSWDLVGRNLFAMA 1857
Db      975 L--LVESAVVEGAMAAVGQSSFAPDYAKAKISAHIIIMIEKPLLDVSTEGSLMNT 1030
QY      1858 VAGVVEFLITVLIQYREFIRP--PYNAKLSPLNDEDEVDREKRIIDGGGNDILEKE 1916
Db      1031 LEGNVTFFGEV--FNYPLRPDIPVLQGLS-----LEVKK 1062
QY      1917 LTKYRRRRKRPVDRICVIGIPGECFGLGVNAGKSTFKMLTQDVTTRGDAFLNKS 1976
Db      1063 -----GQTLALVGSSGCKSTVVOLETFYPLAGKYLDDKE 1100
QY      1977 TILS-NIEVHQMNGYCPQ-----FD-----AITELLTGREHVEFFALLRG 2015
Db      1101 IKRLNVOMLRALHGIIVSOEPIFLDCSIAENIAYGDNRSVSOEIVRAKKAANIHAFTES 1160
QY      2016 VBEKEVGAVGEWAIRKGLVYKGEYAGNYSQGNKRKISTAMALIGPPVYFLDEPTTGM 2075
Db      1161 LFNKYSTKVGDKGTQ-----LSGGQKQRIAIRLVRQPHILLDEAFISAL 1206
QY      2076 DEKARFLMNCALSVYKESRSVLTSHSMECEALCTRMALMVNGREFGLSVOLHKNRF 2135
Db      1207 DESEKVVQGE-ALDKAREGRCTIVIAHLSTIQN-ADLIVFQNGRVAKEHGTQOQLLAQK 1264
QY      2136 GDGYIIVIRIAGS 2148
Db      1265 GIFYSMVSVQACT 1277

RESULT 13
US-09-134-001C-3369
; Sequence 3369, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3369
; LENGTH: 460
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3369

Query Match      2.7%; Score 315.5; DB 4; Length 460;
Best Local Similarity 24.6%; Pred. No. 1.2e-20;
Matches 99; Conservative 87; Mismatches 144; Indels 73; Gaps 13;

QY      899 VSIQNLVYVRNGMKVAVADGLALNTEGQITSLFGLHGNAGKTTMTSLTGLFPPISTGAY 958
Db      44 LSIKMLTITY-SGNKRAVDNISLDIQSGEFLAIGTSGSKTTLARMNRMIEATDQIM 102
QY      959 ILGKDIRS-EMSTIRQNLGVCPQHNVLFDMLTVEEHIMFYRLKGLSKHYKAEQKAL 1017
Db      103 MNGKXVRNNPVELRLRSIGYVIOIGLMPHMTIRENIVLVKLLKMSKEKDEKAKELIK 162
QY      1018 DVGDPSSKIKSTSQLSGMGOKISVALAFVGGKSVVILDEPTAGVDYPSRGJWEILL-- 1075
Db      163 LVDPPEEVLDRYPALBSGGQOQIRIGVRAALAAEODIILMEDEFGLDITRDTIQDLVKE 222

```

```

QY      1076 LKRYGRGRTIILSTHNDADVLDGRIRAIISHGK-----LCCVSSSLFLKNQLGTCY 1126
Db      223 LQKRLCKTFIFVYHMDDEALKLADKICIMSKGVYQYDPDNLIRPAPDFVRDIDGN- 281
QY      1127 YLTLVK-----KDVESLSS-----CNSSSTVSYLKK 1154
Db      282 --RLIDRPNMKSVESAMIKPVYKADDSLINDAVNIMRTRRVDTTFVNNQNKLLGFLDI 339
QY      1155 EDSVSSSSDAGISGDHE--SDTLTIDV-----SAISNLIRKHVEARLVEDIGH 1203
Db      340 ED-INQ-----GLRARELIDTWQRDYVYVHNSKIQDSVFTILKRVNANVEVDDEHL 393
QY      1204 LTYLVPYEAKEGAFVELFHEIDRLSLGSSYGISETTLEE 1246
Db      394 IGLI-----TRANLVDIY--DSIWGEEDSDSYELPNESLDE 428

RESULT 14
US-09-120-513-2
; Sequence 2, Application US/09120513
; Patent No. 6025160
; GENERAL INFORMATION:
; APPLICANT: Brun, Kimberly
; APPLICANT: Chenery, Richard
; APPLICANT: Eliens, Harma
; APPLICANT: Field, John
; APPLICANT: Yue, Lin
; TITLE OF INVENTION: POLYNUCLEOTIDE AND POLYPEPTIDE
; TITLE OF INVENTION: SEQUENCES ENCODING RAT MDR1B2 AND
; TITLE OF INVENTION: SCREENING METHODS THEREOF
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY:
; ZIP: 19406
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/120,513
; FILING DATE: 22-JUL-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: King, William T
; REGISTRATION NUMBER: 30,954
; REFERENCE/DOCKET NUMBER: GP50008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5015
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1275 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-120-513-2

Query Match      2.6%; Score 311; DB 3; Length 1275;
Best Local Similarity 19.0%; Pred. No. 2.6e-19;
Matches 281; Conservative 180; Mismatches 437; Indels 580; Gaps 62;

QY      743 ANLAACGGITTYTLVLPVLCVAMQDYGFTLKIFASLSFYAGSGCGCFALFEQGI 802
Db      293 ANSI--GIAVLVYASVAL-----AFWYGN-SLVLSNEYSI 326

```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 7, 2003, 08:41:19 ; Search time 41 seconds
(without alignments)
2287.268 Million cell updates/sec

Title: US-09-595-526C-2

Perfect score: 11797

Sequence: 1 MACWPOLRLMLKMLTFRFR.....VDVALTSLFODEKVEKESYV 2261

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

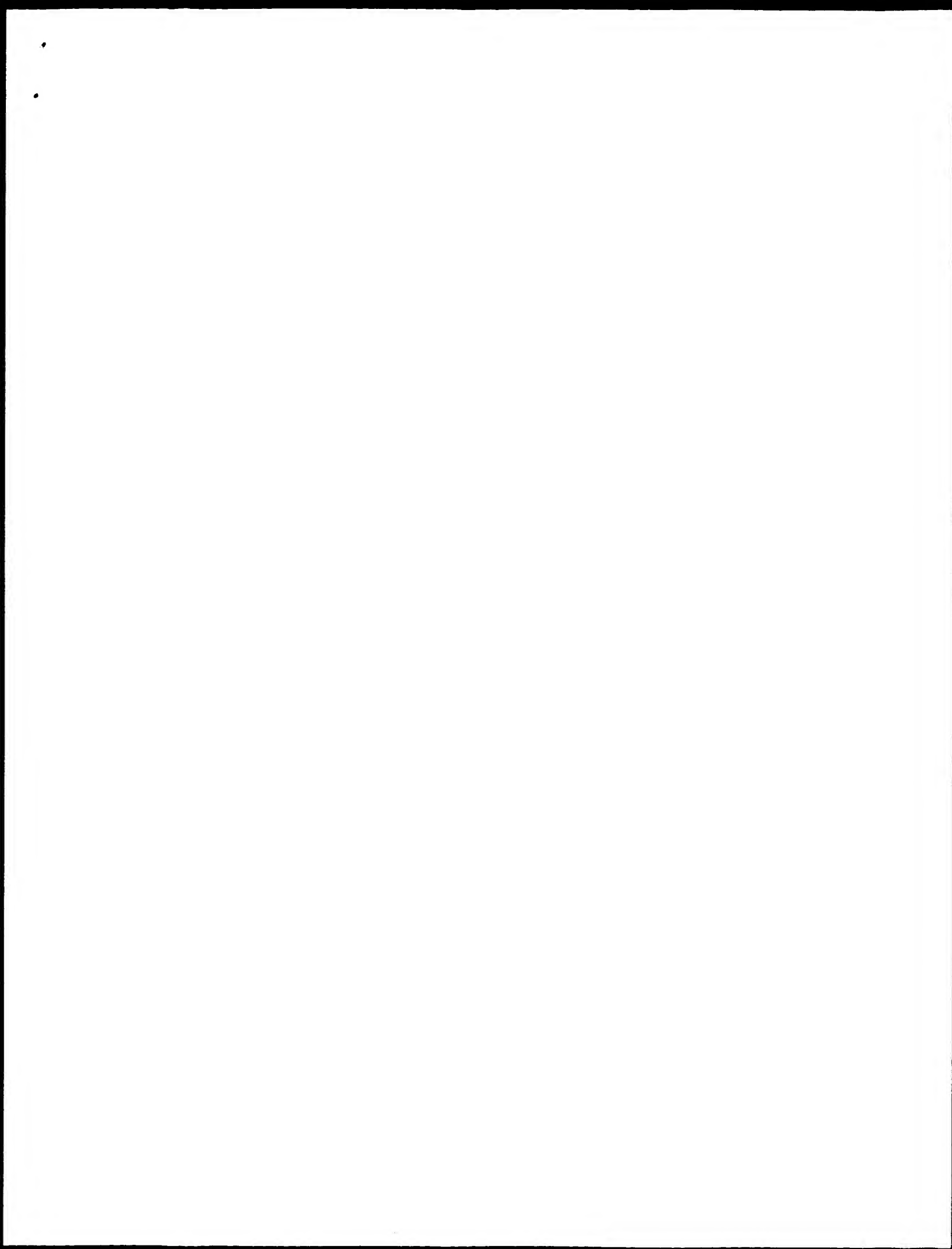
SUMMARIES

Result No.	Score	Query	Length	ID	Description
1	11770	99.8	2261	1	ABCI_HUMAN
2	11256	95.4	2261	1	ABCI_MOUSE
3	5870.5	49.8	2273	1	ABCR_HUMAN
4	4230.5	35.9	2436	1	ABCR2_HUMAN
5	4093.5	34.7	2434	1	ABCR2_MOUSE
6	2642.5	22.4	1704	1	ABCI_HUMAN
7	1538.5	13.0	1704	1	CBP7_CAEEL
8	411	3.5	330	1	DRRA-STRE
9	382.5	3.2	343	1	NOD1_RHISN
10	366	3.1	304	1	NOD1_RHIS3
11	347	2.9	308	1	YADG_ECOLI
12	344.5	2.9	340	1	NOD1_RHIC
13	343.5	2.9	347	1	NOD1_RHIC
14	339.5	2.9	1280	1	MDR1_HUMAN
15	331.5	2.8	355	1	NOD1_RHIME
16	329.5	2.8	578	1	YBHF_ECOLI
17	327.5	2.8	1281	1	MDR3_CRIGR
18	327	2.8	894	1	YH1H_ECOLI
19	326	2.8	1276	1	MDR3_MOUSE
20	325.5	2.8	308	1	NOSF_PEST
21	324.5	2.8	1276	1	MDR2_MOUSE
22	322.5	2.7	306	1	NOD1_BRAJA
23	321	2.7	1276	1	MDR1_CRIGR
24	317	2.7	1362	1	PMO1_SCHPO
25	316	2.7	1276	1	MDR2_CRIGR
26	315.5	2.7	311	1	NOD1_RHILV
27	315	2.7	1278	1	MDR2_RAT
28	314.5	2.7	1277	1	MDR1_RAT
29	313	2.7	354	1	Y415_SYNY3
30	310.5	2.6	1276	1	MDR1_MOUSE
31	304.5	2.6	1284	1	YOH5_YEAST
32	300	2.5	381	1	OPBA_BACSU
33	298.5	2.5	262	1	YA23_METJA

34	297	2.5	380	1	OPCA_BACSU
35	296	2.5	335	1	Y719_ANASP
36	293.5	2.5	1321	1	AB11_HUMAN
37	292.5	2.5	305	1	YHCH_BACSU
38	287.5	2.4	274	1	Y179_MYCPN
39	286.5	2.4	274	1	Y179_MYCGE
40	285.5	2.4	1279	1	MDR3_HUMAN
41	284	2.4	306	1	BCRA_BACLI
42	284	2.4	1321	1	MDR1_CAEEL
43	284	2.4	1336	1	MAMI_SCHPO
44	281.5	2.4	308	1	YEHX_ECOLI
45	281.5	2.4	343	1	ABCI_ECOLI

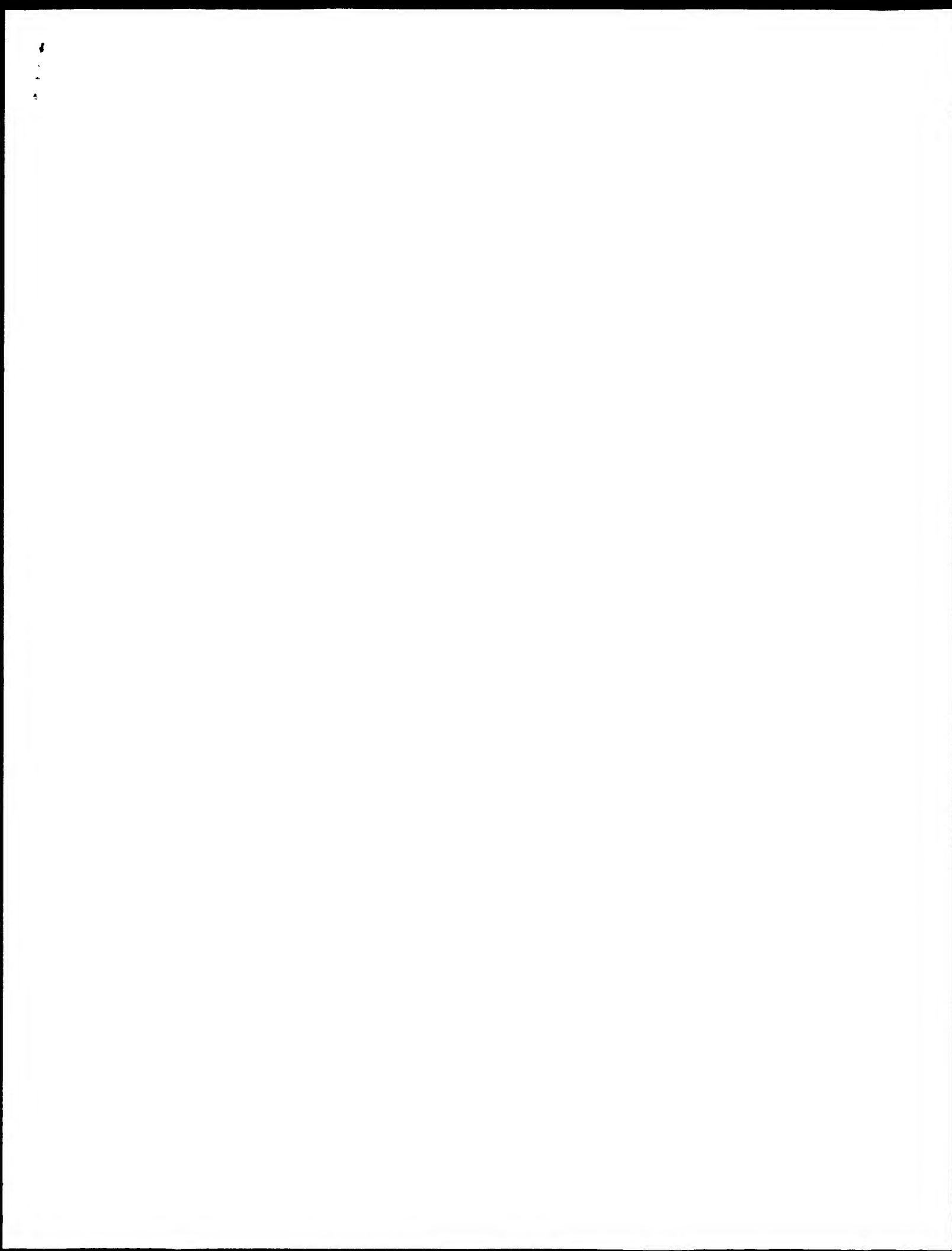
ALIGNMENTS

RESULT 1
ABCI_HUMAN STANDARD; PRT; 2261 AA.
AC 095477; Q9UN08; Q9UN07; Q9UN06; Q9UN09; Q96785; Q96556;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ATP-binding cassette, sub-family A, member 1 (ATP-binding cassette transporter 1) (ATP-binding cassette 1) (ABC-1) (Cholesterol efflux regulatory protein).
DE ABCA1 OR ABC1 OR CERP.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_TaxID:9606;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE-20345099; PubMed-10884428;
RA Santamarina-Fojo S., Peterson K.M., Knapper C.L., Qiu Y., Freeman L.A., Cheng J.-F., Osorio J., Remaley A.T., Yang X.-P., Haudenschield C.C., Prades C., Chimini G., Blackmon E.E., Francois T.L., Duvenger N., Rubin E.M., Rosier M., Deneffe P., Fredrickson D.S., Brewer H.B. Jr.;
RA "Complete genomic sequence of the human ABCA1 gene: analysis of the human and mouse ATP-binding cassette A promoter.";
RT Proc. Natl. Acad. Sci. U.S.A. 97:7987-7992(2000). -544.
[2]
RP TISSUE-SKIN; N.A.
RA Schwartz K., Lavin R.M., Wade D.P.;
RT "ABCA1 gene expression and apoA-I-mediated cholesterol efflux are regulated by LXR.";
RN Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RX MEDLINE-21251004; PubMed-11352567;
RA Qiu Y., Caveller L., Chiu S., Yang X., Rubin E., Cheng J.-F.;
RT "Human and mouse ABCA1 comparative sequencing and transgenesis studies revealing novel regulatory sequences.";
RN Genomics 73:66-76(2001).
[4]
RP SEQUENCE FROM N.A.
RA Tanaka A.R., Abe-Dohmae S., Arakawa R., Sadaami K., Kidera A., Kikita N., Amachi T., Yokoyama S., Ueda K.;
RT "A new topological model of functional human ABCA1-signal peptide cleavage and glycosylation of a large extracellular domain.";
RN Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
[5]
RP SEQUENCE OF 21-2261 FROM N.A.
RX MEDLINE-99194549; PubMed-10092505;
RA Langman T., Klucken J., Reil M., Liebisch G., Luciani M.F., Chimini G., Kaminski W.E., Schmitz G.;
RT "Molecular cloning of the human ATP-binding cassette transporter 1 (ABCA1): evidence for sterol-dependent regulation in macrophages.";
RN Biochem. Biophys. Res. Commun. 257:29-33(1999). [6]



RP SEQUENCE OF 21-2261 FROM N.A.
 RX MEDLINE=99364413; PubMed=10431238;
 RA Rust S., Rosier M., Funke H., Real J., Amoura Z., Piette J.-C.,
 RA Deleuze J.-F., Brewer H.B., Duverger N., Deneffe P., Assmann G.;
 RT "Tangier disease is caused by mutations in the gene encoding
 RT ATP-binding cassette transporter 1.";
 RL Nat. Genet. 22:352-355(1999).
 RN [17]
 RP VARIANTS FHA THR-1091 AND 1893-GLU-ASP-1894 DEL.
 RX MEDLINE=20001430; PubMed=10533863;
 RA Marcell M., Brooks-Wilson A., Clee S.M., Roomp K., Zhang L.-H., Yu L.,
 RA Collins J.A., van Dam M., Molhuizen H.O.F., Loubser O.,
 RA Ouellette B.F.F., Sensen C.W., Fichter K., Mott S., Denis M.,
 RA Boucher B., Pimstone S., Genest J. Jr., Kastelein J.J.P., Hayden M.R.;
 RT "Mutations in the ABCA1 gene in familial HDL deficiency with defective
 RT cholesterol efflux.";
 RL Lancet 354:1341-1346(1999).
 RN [8]
 RP VARIANTS TD ARG-597 AND ARG-1477, AND VARIANT FHA LEU-693 DEL.
 RX MEDLINE=99364411; PubMed=10431236;
 RA Brooks-Wilson A., Marcell M., Clee S.M., Zhang L.-H., Roomp K.,
 RA van Dam M., Yu L., Brewer C., Collins J.A., Molhuizen H.O.F.,
 RA Loubser O., Ouellette B.F.F., Fichter K., Ashbourne-Excoffon K.J.D.,
 RA Sensen C.W., Scherer S., Mott S., Denis M., Martindale D.,
 RA Frohlich J., Morgan K., Koop B., Pimstone S., Kastelein J.J.P.,
 RA Hayden M.R.;
 RT "Mutations in ABCA1 in Tangier disease and familial high-density
 RT lipoprotein deficiency.";
 RL Nat. Genet. 22:336-345(1999).
 RN [9]
 RP VARIANTS TD SER-590; SER-935 AND VAL-937, AND VARIANTS ALA-399 AND
 RP MET-883
 RX MEDLINE=99364412; PubMed=10431237;
 RA Bodzioch M., Orso E., Klucken J., Langmann T., Boettcher A.,
 RA Diederich W., Drobnik W., Barlage S., Buechler C.,
 RA Porsche-Oezcuerumez M., Kaminski W.E., Hahmann H.W., Oette K.,
 RA Rothe G., Aslanidis C., Lackner K.J., Schmitz G.;
 RT "The gene encoding ATP-binding cassette transporter 1 is mutated in
 RT Tangier disease.";
 RL Nat. Genet. 22:347-351(1999).
 RN [10]
 RP VARIANTS TD ILE-929; ARG-597 AND ARG-1477, AND VARIANTS FHA LEU-693
 RP DEL; THR-1091; 1893-GLU-ASP-1894 DEL AND LEU-2150.
 RX MEDLINE=20540002; PubMed=11086027;
 RA Clee S.M., Kastelein J.J.P., van Dam M., Marcell M., Roomp K.,
 RA Zwarts K.Y., Collins J.A., Roelants R., Tamawana N., Stult T.,
 RA Suda T., Ceska R., Boucher B., Rondeau C., Desouch C.,
 RA Brooks-Wilson A., Molhuizen H.O.F., Frohlich J., Genest J. Jr.,
 RA Hayden M.R.;
 RT "Age and residual cholesterol efflux affect HDL cholesterol levels and
 RT coronary artery disease in ABCA1 heterozygotes.";
 RL J. Clin. Invest. 106:1263-1270(2000).
 RN [11]
 RP VARIANTS TD ASN-1289 AND HIS-1800.
 RX MEDLINE=20171564; PubMed=10706591;
 RA Brousseau M.E., Schaefer E.J., Dupuis J., Eustace B.,
 RA Van Berdevegh P., Goldkamp A.L., Thurston L.M., Fitzgerald M.G.,
 RA Yasek-McKenna D., O'Neill G., Eberhart G.P., Wellfendach B.,
 RA Orlovsk J.M., Freeman M.W., Brown R.H. Jr., Gu J.Z.;
 RT "Novel mutations in the gene encoding ATP-binding cassette 1 in four
 RT Tangier disease kindreds.";
 RL J. Lipid Res. 41:433-441(2000).
 RN [12]
 RP VARIANT TD ASP-1046, VARIANT FHA CYS-230, AND VARIANTS LYS-219;
 RP ILE-825; MET-883 AND LYS-1587.
 RX MEDLINE=20396633; PubMed=10938021;
 RA Wang J., Burnett J.R., Near S., Young K., Zimman B., Hanley A.J.G.,
 RA Connelly P.W., Harris S.B., Hegde R.A.;
 RT "Common and rare ABCA1 variants affecting plasma HDL cholesterol.";
 RL Arterioscler. Thromb. Vasc. Biol. 20:1963-1969(2000).
 RN [13]
 RP VARIANT TD TRP-587, AND VARIANT LEU-2168.
 RX MEDLINE=21157002; PubMed=11257260;

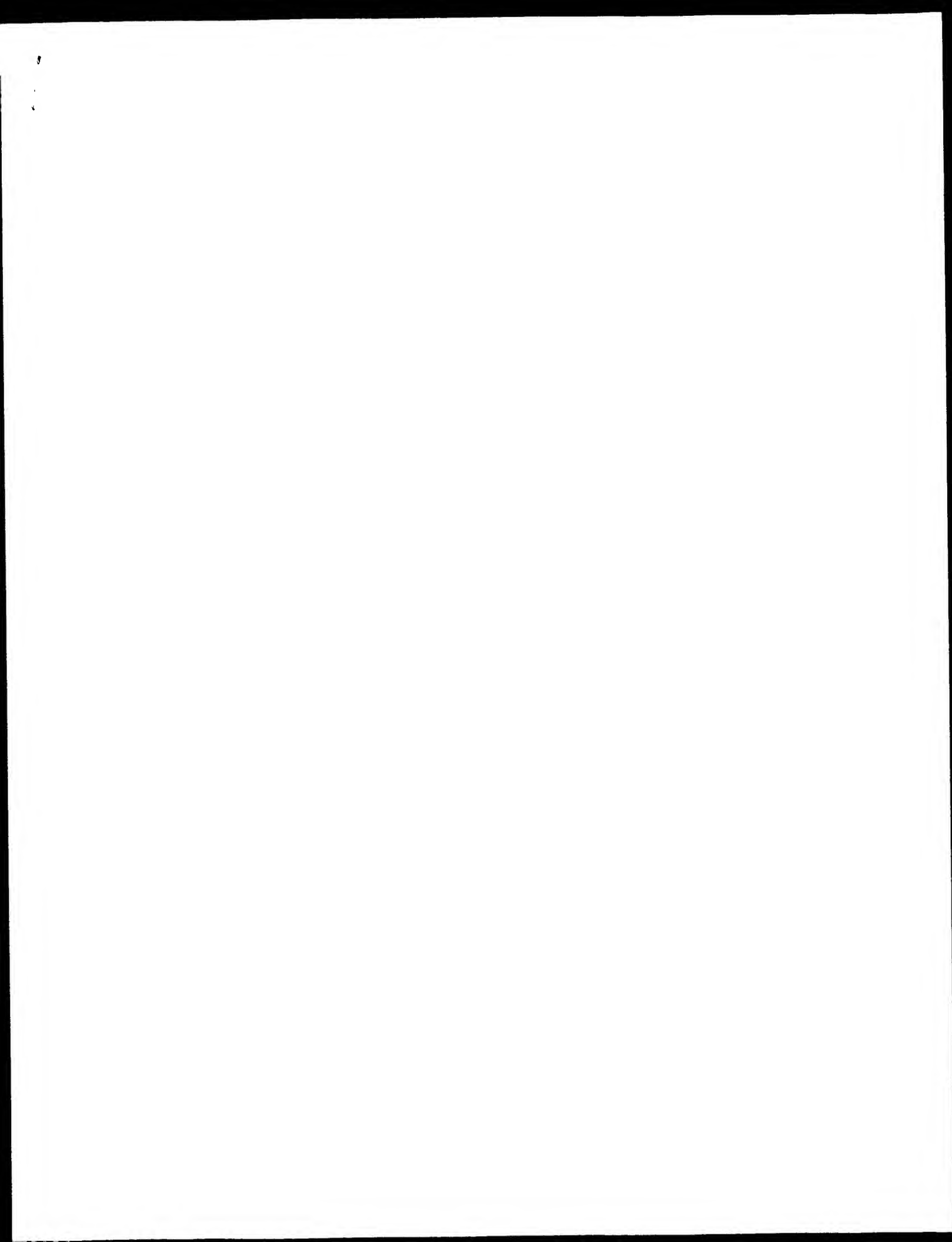
RA Bertolini S., Pisciotto L., Seri M., Cusano R., Cantalora A.,
 RA Calabresi L., Franceschini G., Ravazzolo R., Calandra S.;
 RT "A point mutation in ABCA1 gene in a patient with severe premature
 RT coronary heart disease and mild clinical phenotype of Tangier
 RT disease.";
 RL Atherosclerosis 154:599-605(2001).
 RN [14]
 RP VARIANTS LYS-219; MET-883 AND ASP-1172.
 RX MEDLINE=21157003; PubMed=11257261;
 RA Brousseau M.E., Bodzioch M., Schaefer E.J., Goldkamp A.L., Kielar D.,
 RA Probst M., Orlovsk J.M., Aslanidis C., Lackner K.J.,
 RA Bloomfield Rubins H., Collins D., Robins S.J., Wilson P.W.F.,
 RA Schmitz G.;
 RT "Common variants in the gene encoding ATP-binding cassette transporter
 RT 1 in men with low HDL cholesterol levels and coronary heart disease.";
 RL Atherosclerosis 154:607-611(2001).
 RN [15]
 RP VARIANT TD LEU-1506.
 RX MEDLINE=21369429; PubMed=11476961;
 RA Lapack-Bodzioch K., Bodzioch M., Kruehl M., Kielar D., Probst M.,
 RA Kiec B., Andrikovics H., Boettcher A., Hubacek J., Aslanidis C.,
 RA Suttorp N., Schmitz G.;
 RT "Homogeneous assay based on 52 primer sets to scan for mutations of
 RT the ABCA1 gene and its application in genetic analysis of a new
 RT patient with familial high-density lipoprotein deficiency syndrome.";
 RL Biochim. Biophys. Acta 1537:42-48(2001).
 RN [16]
 RP VARIANTS TD ASN-1289 AND TRP-2081, AND VARIANT LYS-219.
 RX MEDLINE=21369433; PubMed=11476965;
 RA Huang W., Moriyama K., Koga T., Hua H., Ageta M., Kawabata S.,
 RA Nawatari K., Imanura T., Eto T., Kawamura M., Teramoto T., Sasaki J.;
 RT "Novel mutations in ABCA1 gene in Japanese patients with Tangier
 RT disease and familial high density lipoprotein deficiency with
 RT coronary heart disease.";
 RL Biochim. Biophys. Acta 1537:71-78(2001).
 RN [17]
 RP VARIANTS LYS-219; ALA-399; MET-771; PRO-774; ASN-776; ILE-825;
 RP MET-883; ASP-1172; LYS-1587 AND CYS-1731.
 RX MEDLINE=2138379; PubMed=11238261;
 RA Clee S.M., Zwinderman A.H., Engert J.C., Zwarts K.Y.,
 RA Molhuizen H.O.F., Roomp K., Jukema J.W., van Wijland M., van Dam M.,
 RA Hudson T.J., Brooks-Wilson A., Genest J. Jr., Kastelein J.J.P.,
 RA Hayden M.R.;
 RT "Common genetic variation in ABCA1 is associated with altered
 RT lipoprotein levels and a modified risk for coronary artery disease.";
 RL Circulation 106:1198-1205(2001).
 RN [18]
 RP VARIANT TD THR-255, AND VARIANT ATHEROSCLEROSIS ASP-1611,
 RX MEDLINE=21645894; PubMed=11785958;
 RA Nishida Y., Hirano K., Tsukamoto K., Nagano M., Ikegami C., Roomp K.,
 RA Ishihara M., Sakane N., Zhang Z., Tsujii K., Matsuyama A., Ohama T.,
 RA Matsuda F., Ishigami M., Sakai N., Hirooka H., Hattori H.,
 RA Yamashita S., Matsuzawa Y.;
 RT "Expression and functional analyses of novel mutations of ATP-binding
 RT cassette transporter-1 in Japanese patients with high-density
 RT lipoprotein deficiency.";
 RL Biochem. Biophys. Res. Commun. 290:713-721(2002).
 CC -1- FUNCTION: CAMP-DEPENDENT AND SUBUNITURKA-SENSITIVE ANION
 CC TRANSPORT. KEY GATEKEEPER INFLUENCING INTRACELLULAR CHOLESTEROL
 CC TRANSPORT.
 CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED, BUT MOST ABUNDANT IN
 CC MACROPHAGES.
 CC -1- DOMAIN: MULTIFUNCTIONAL POLYPEPTIDE WITH TWO HOMOLOGOUS HALVES,
 CC EACH CONTAINING AN HYDROPHOBIC MEMBRANE-ANCHORING DOMAIN AND AN
 CC ATP BINDING CASSETTE (ABC) DOMAIN.
 CC -1- DISEASE: DEFECTS IN ABCA1 ARE A CAUSE OF HIGH DENSITY LIPOPROTEIN
 CC DEFICIENCY TYPE I (HDLI), ALSO KNOWN AS TANGIER DISEASE (TD). TD
 CC IS A RECESSIVE DISORDER CHARACTERIZED BY ABSENCE OF HIGH DENSITY
 CC LIPOPROTEIN (HDL) CHOLESTEROL FROM PLASMA, HEPATOSPLENOMEGALY,
 CC PERIPHERAL NEUROPATHY, AND FREQUENTLY PREMATURE CORONARY ARTERY
 CC DISEASE (CAD).
 CC -1- DISEASE: Defects in ABCA1 are a cause of high density lipoprotein



Query Match 99.8%; Score 11770; DB 1; Length 2261;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 2255; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

1 MACWQRLRLMKNTLFRRCOTCOLLEAVWPLFLILISVRLSPYEQHECHFPNKA 60
1 MACWQRLRLMKNTLFRRCOTCOLLEAVWPLFLILISVRLSPYEQHECHFPNKA 60
61 MSAGTLPWVQIICNANNPCFRYPGPAGVGNFNKSIYARLFSPARLLYSQDT 120
61 MSAGTLPWVQIICNANNPCFRYPGPAGVGNFNKSIYARLFSPARLLYSQDT 120
121 SKKDKRKLVTLOQIKKSSNLKLODFLVNTEFSGLTYHNLSPKSTYDKMLRADVILH 180
121 SKKDKRKLVTLOQIKKSSNLKLODFLVNTEFSGLTYHNLSPKSTYDKMLRADVILH 180
181 KYFLQGYOLHLSLNGSKSEPMIOGQOEVSELCGLPREKILAAERVLRSNMDILKPTL 240
181 KYFLQGYOLHLSLNGSKSEPMIOGQOEVSELCGLPREKILAAERVLRSNMDILKPTL 240
241 RLINSTSPSPSKELAEATKTLHSLGTLAQLFMSRSMDSMRQEVMTLNNSSSSSTQI 300
241 RLINSTSPSPSKELAEATKTLHSLGTLAQLFMSRSMDSMRQEVMTLNNSSSSSTQI 300
301 YQAVSRIVCGHPEGGLKIKSLNWTEDNNYKALFEGNGTEDEAETFYDNSTTPYCNDLMK 360
301 YQAVSRIVCGHPEGGLKIKSLNWTEDNNYKALFEGNGTEDEAETFYDNSTTPYCNDLMK 360
361 NLESPSLRIIMKALKPLVYKIIYTPPTPATROYMAEVNTPQELAVFHLEGMWELLS 420
361 NLESPSLRIIMKALKPLVYKIIYTPPTPATROYMAEVNTPQELAVFHLEGMWELLS 420
421 PKIWTFMNSQBDLVRLDLSDRDHFWEQDGLDWTADIVAFLLAKHPEDVOSSNGS 480
421 PKIWTFMNSQBDLVRLDLSDRDHFWEQDGLDWTADIVAFLLAKHPEDVOSSNGS 480
481 VYTRAEAFENETNOAIRTSRMECVNLKLEPIATEVWLINXSHLDERFNAGIYFTG 540
481 VYTRAEAFENETNOAIRTSRMECVNLKLEPIATEVWLINXSHLDERFNAGIYFTG 540
541 ITPPSIELPHHVKKYKIRMDIDNVERTNKIKDGYMDPGPRADPFDMRYWGCFAIYLDVY 600
541 ITPPSIELPHHVKKYKIRMDIDNVERTNKIKDGYMDPGPRADPFDMRYWGCFAIYLDVY 600
601 EOLAIRVLTEGKRTGYVMQMPYPCYVDIIFLVKSRMSMLEMTLMTYSVAVIKGIY 660
601 EOLAIRVLTEGKRTGYVMQMPYPCYVDIIFLVKSRMSMLEMTLMTYSVAVIKGIY 660
661 YEKARLKEETMRINGLDNSILMFEMTSSILPLIVSAGLLVYLKLGNLPLPSDPVYEV 720
661 YEKARLKEETMRINGLDNSILMFEMTSSILPLIVSAGLLVYLKLGNLPLPSDPVYEV 720
721 FLSPFAVVTIIQCFILISTLFSRANLAAACGIIYFTLYLPVLCVAMQDYVGFLLKFTAS 780
721 FLSPFAVVTIIQCFILISTLFSRANLAAACGIIYFTLYLPVLCVAMQDYVGFLLKFTAS 780
781 LLSVAFGFGCEYALFEEDGIGYQWNLPESPVEEDGFNLITTSISMLFDFELVGYMTW 840
781 LLSVAFGFGCEYALFEEDGIGYQWNLPESPVEEDGFNLITTSISMLFDFELVGYMTW 840
841 YIEAVPFGOYIGIPRPWFYFCTKSYWFGESDESKHPSGNOKRMSJICMEDEPHLIKIGVS 900
841 YIEAVPFGOYIGIPRPWFYFCTKSYWFGESDESKHPSGNOKRMSJICMEDEPHLIKIGVS 900
901 IONLVKYYRGMKAVADGLALNFYEQITSFLGHNGAKTTTMSILTLGLPPTSGATYIL 960
901 IONLVKYYRGMKAVADGLALNFYEQITSFLGHNGAKTTTMSILTLGLPPTSGATYIL 960
961 GKIRSEMSITRONLIGVCPQHNVLFLDKLTYEEHIMFARLKGISEKHVAKEMOMALDVG 1020
961 GKIRSEMSITRONLIGVCPQHNVLFLDKLTYEEHIMFARLKGISEKHVAKEMOMALDVG 1020

1021 LPSSKIKSKTSQLSGQORLVALALAVGSKYVILDEPPAGVDPPYSRRGIWELLKRYQ 1080
1021 LPSSKIKSKTSQLSGQORLVALALAVGSKYVILDEPPAGVDPPYSRRGIWELLKRYQ 1080
1081 GRTIILSTHMHDEADVLDGDIATIIISHGKLCCVSSSFLKNOLOETGYVLTLYKKDVSSIS 1140
1081 GRTIILSTHMHDEADVLDGDIATIIISHGKLCCVSSSFLKNOLOETGYVLTLYKKDVSSIS 1140
1141 SCRNSSTVSYLLKEDSVSSSSDAGLGSDESDTLITIDVSAISNLRKHVSARLVEDI 1200
1141 SCRNSSTVSYLLKEDSVSSSSDAGLGSDESDTLITIDVSAISNLRKHVSARLVEDI 1200
1201 GHELTLYLPEAKKEGAFVLFHEIDRLSDLGISYGISSETTLEELFLKVAEESVDAE 1260
1201 GHELTLYLPEAKKEGAFVLFHEIDRLSDLGISYGISSETTLEELFLKVAEESVDAE 1260
1261 TSDGTLPARNRRAFGDKQCLRPTEEDADPNDSIDIPESRETFLLSGMDKGSYQK 1320
1261 TSDGTLPARNRRAFGDKQCLRPTEEDADPNDSIDIPESRETFLLSGMDKGSYQK 1320
1321 GWKLTQOQFVALLMKRLILARSRKGFPAQIYLPVAVFCIALVSLIIPPCKYPSLEQ 1380
1321 GWKLTQOQFVALLMKRLILARSRKGFPAQIYLPVAVFCIALVSLIIPPCKYPSLEQ 1380
1381 PMWYNEQYTVSNDAPDGTGTLFLNALTKDPGEGRKQMBGNIPDTPQAGEEETTPAP 1440
1381 PMWYNEQYTVSNDAPDGTGTLFLNALTKDPGEGRKQMBGNIPDTPQAGEEETTPAP 1440
1441 VPOTIMDLFONGWMTQONPSPACQSSDKIKKMLPVCPPGAGGLPPQKQONTADILQDL 1500
1441 VPOTIMDLFONGWMTQONPSPACQSSDKIKKMLPVCPPGAGGLPPQKQONTADILQDL 1500
1501 TGRNISDYIKTYVOIILASIKKXIVWNEFRYGGESLVSNTQALPPSOEVDATKQMK 1560
1501 TGRNISDYIKTYVOIILASIKKXIVWNEFRYGGESLVSNTQALPPSOEVDATKQMK 1560
1561 HLKLANDSSADRELNSLGRFMTGLDTRNNVKYWFENKKGHAISFLVNTNNAILRANLQK 1620
1561 HLKLANDSSADRELNSLGRFMTGLDTRNNVKYWFENKKGHAISFLVNTNNAILRANLQK 1620
1621 GENPSHYGITAFAHPLNLTROOLSEVALMTTSYDVLSVTCVIFAMSPVASEFVLEIQR 1680
1621 GENPSHYGITAFAHPLNLTROOLSEVALMTTSYDVLSVTCVIFAMSPVASEFVLEIQR 1680
1681 VSKAKHLQFISGVKPYIYMLSNFVMDKNTVVPATLVIIIFICFOOKSVYSTNLPLYAL 1740
1681 VSKAKHLQFISGVKPYIYMLSNFVMDKNTVVPATLVIIIFICFOOKSVYSTNLPLYAL 1740
1741 LLLLYGMSITPLMPYASFEFKIPSTAYVVLTSVNLFIGINGSVAFVLELFTDNKLNIN 1800
1741 LLLLYGMSITPLMPYASFEFKIPSTAYVVLTSVNLFIGINGSVAFVLELFTDNKLNIN 1800
1801 DILKSVFLIFPHCLGRGLIDMKYKNOAMADALERRGENFVSPLSMDLVGRNLFMAVEG 1860
1801 DILKSVFLIFPHCLGRGLIDMKYKNOAMADALERRGENFVSPLSMDLVGRNLFMAVEG 1860
1861 VVEFLTIVLIQYREFIRPPVNAKLSPLNDEDEDEYRREORLILDGGGNDILIEIKELKI 1920
1861 VVEFLTIVLIQYREFIRPPVNAKLSPLNDEDEDEYRREORLILDGGGNDILIEIKELKI 1920
1921 YRRKRPAYDRICVGIIPGECFGLLVGNAGKSSFFKMLTGDYTVTRDAPLANKSISLN 1980
1921 YRRKRPAYDRICVGIIPGECFGLLVGNAGKSSFFKMLTGDYTVTRDAPLANKSISLN 1980
1981 IHEVHONMGYCPQPAITELLGHEHVEFFALLRCVPEKXGVGEAIIKRLGLVYKEX 2040
1981 IHEVHONMGYCPQPAITELLGHEHVEFFALLRCVPEKXGVGEAIIKRLGLVYKEX 2040
2041 YAGNTSGGNKRLSTAMALIGGPVYVFLDEPTTGMDPARPRLMICALSVYKEGSSVLT 2100
2041 YAGNTSGGNKRLSTAMALIGGPVYVFLDEPTTGMDPARPRLMICALSVYKEGSSVLT 2100
2101 SHSMECEALCTRNAIMVNGRFLGSGVOHLKNRFQGDYTVTVRIAGSNPDLKPVODFFG 2160




```

Db 2101 SHSMECEALCTRAIMVNGFRCLGSVQHLKNRFGDYITLVIRAGSNPLKVPQDFEG 2160
Oy 2161 LAFPSVLAKEKRRMLQYQLPSSLSLARIRISLSQSKRRILHIEDYVSQTTLDQVYNF 2220
Db 2161 LAFPSVLAKEKRRMLQYQLPSSLSLARIRISLSQSKRRILHIEDYVSQTTLDQVYNF 2220
Oy 2221 AKDQSDHKLKDLHLKNOYVAVLFSPLQDEKVKESYV 2261
Db 2221 AKDQSDHKLKDLHLKNOYVAVLFSPLQDEKVKESYV 2261

RESULT 2
ID ABCL_MOUSE STANDARD: PRT: 2261 AA.
AC P41233.
DT 01-FEB-1995 (Rel. 31, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ATP-binding cassette, sub-family A, member 1 (ATP-binding cassette transporter 1) (ATP-binding cassette 1) (ABC-1).
GN ABCA1 OR ABC1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DKA/2; TISSUE=Macrophage;
RX MEDLINE=94375008; PubMed=8088782;
RA Luciani M.F., Denizot F., Savary S., Mattei M.-G., Chimi G.;
RT "Cloning of two novel ABC transporters mapping on human chromosome 9".
RL Genomics 21:150-159(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RA Qiu Y., Cavelier L., Chiu S., Rubin E., Cheng J.-F.;
RT "Human and mouse ABCA1 comparative sequencing and transgenesis studies identify potential regulatory sequences."
RL Submitted (Jul-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CAMP-DEPENDENT AND SULFONYLUREA-SENSITIVE ANION TRANSPORTER. KEY GATEKEEPER INFLUENCING INTRACELLULAR CHOLESTEROL TRANSPORT (BY SIMILARITY).
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED IN ADULT TISSUES. HIGHEST LEVELS ARE FOUND IN PREGNANT UTERUS.
CC -1- DOMAIN: MULTIFUNCTIONAL POLYPEPTIDE WITH TWO HOMOLOGOUS HALVES, EACH CONTAINING AN HYDROPHOBIC MEMBRANE-ANCHORING DOMAIN AND AN ATP BINDING CASSETTE (ABC) DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCA SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).
CC
CC EMBL: X75926; CAA53530.1; ALT_INIT.
DR EMBL: AF287263; AAG39073.1; ALT_INIT.
DR MGI: 99607; Abcal.
DR InterPro: IPR003593; AAA_Arpase.
DR InterPro: IPR003439; ABC_transporter.
DR Pfam: PF00005; ABC_tran. 2.
DR ProDom: PD000006; ABC_transportr; 2.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW ATP-binding; Glycoprotein; Transmembrane; Transport.
FT TRANSMEM 26 42 POTENTIAL.
FT TRANSMEM 640 656 POTENTIAL.
FT TRANSMEM 690 706 POTENTIAL.
FT TRANSMEM 717 733 POTENTIAL.

```

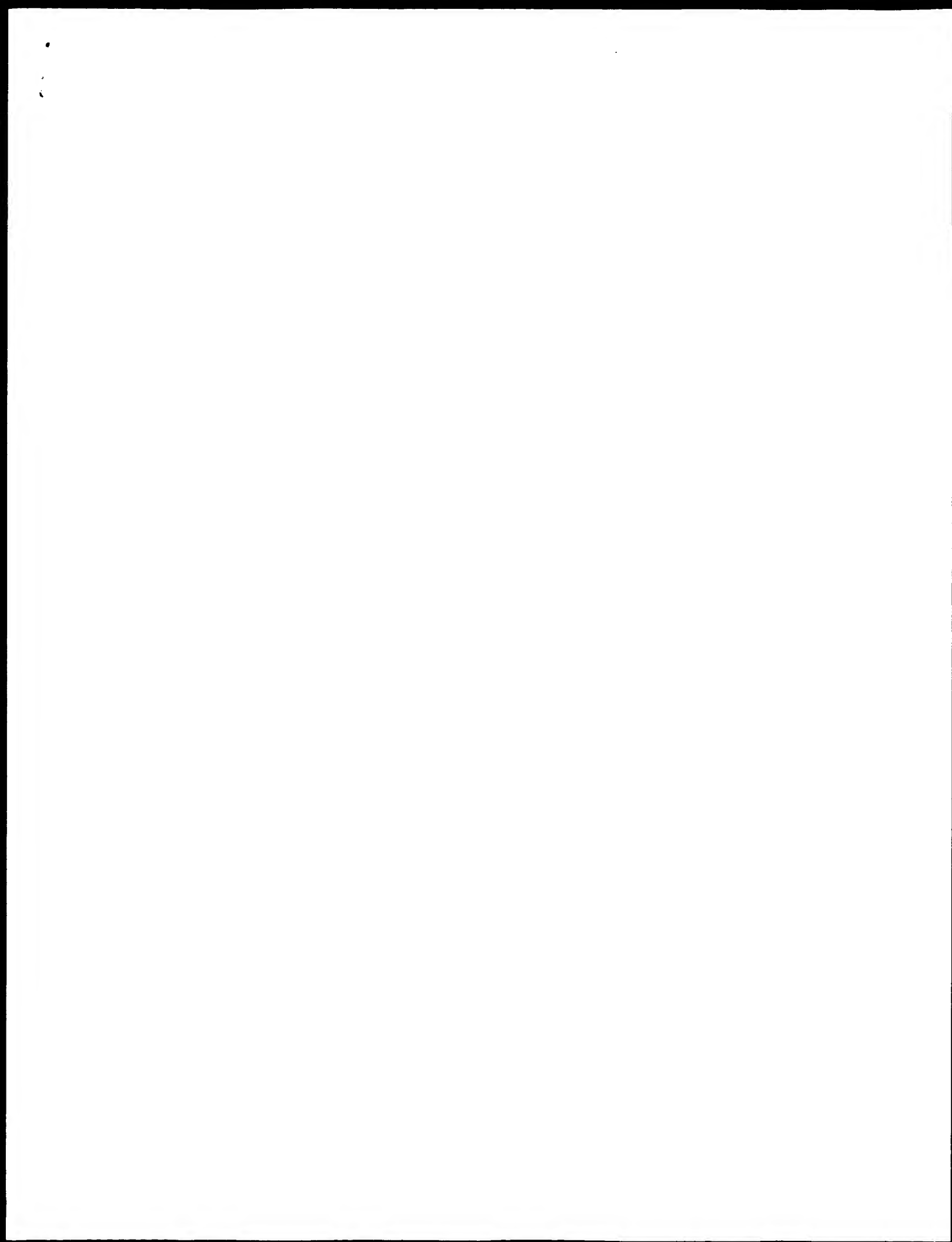
```

FT TRANSMEM 749 765 POTENTIAL.
FT TRANSMEM 771 787 POTENTIAL.
FT TRANSMEM 1041 1057 POTENTIAL.
FT TRANSMEM 1351 1367 POTENTIAL.
FT TRANSMEM 1661 1677 POTENTIAL.
FT TRANSMEM 1708 1724 POTENTIAL.
FT TRANSMEM 1737 1753 POTENTIAL.
FT TRANSMEM 1775 1791 POTENTIAL.
FT TRANSMEM 1854 1870 POTENTIAL.
FT NP_BIND 933 940 ATP (POTENTIAL).
FT NP_BIND 1946 1953 ATP (POTENTIAL).
FT CARBOHYD 14 14 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 98 98 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 151 151 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 161 161 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 244 244 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 292 292 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 337 337 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 349 349 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 400 400 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 478 478 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 489 489 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 521 521 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 820 820 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1144 1144 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1294 1294 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1453 1453 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1499 1499 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1504 1504 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1637 1637 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2044 2044 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2238 2238 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 1567 1568 MISSING (IN REF. 2).
FT CONFLICT 2024 2024 MISSING (IN REF. 2).
SQ SEQUENCE 2261 AA; 254011 MW; FA6E2B21FDD09F96 CRC64;

Query Match 95.4%; Score 11256; DB 1; Length 2261;
Best Local Similarity 95.0%; Pred. No. 0;
Matches 2148; Conservative 55; Mismatches 57; Indels 0; Gaps 0;

Oy 1 MACPOLRLMLKMLNFRROTCOLLVAMPFLFILLISVRLSPYVQHCHEPNKA 60
Db 1 MACPOLRLMLKMLNFRROTCOLLVAMPFLFILLISVRLSPYVQHCHEPNKA 60
Oy 61 MPSAGTLPWVGIIICNNANPCFRPTPEGEAPGVYGNFKSIVARLSDSARLLYSOKRT 120
Db 61 MPSAGTLPWVGIIICNNANPCFRPTPEGEAPGVYGNFKSIVARLSDSARLLYSOKRT 120
Oy 121 SMKMRVRLPTLOQIKKSSNLKLODFLVNMFPSGFLVNNLSLPKSTVYMKRAVYTH 180
Db 121 SMKMRVRLPTLOQIKKSSNLKLODFLVNMFPSGFLVNNLSLPKSTVYMKRAVYTH 180
Oy 181 KVFLOGYQLHTLSJCNCSKSEEMITOLDQOVSSELCGLPKKRLAAERVLRSNMIDILPIL 240
Db 181 KVFLOGYQLHTLSJCNCSKSEEMITOLDQOVSSELCGLPKKRLAAERVLRSNMIDILPIL 240
Oy 181 KVFLOGYQLHTLSJCNCSKSEEMITOLDQOVSSELCGLPKKRLAAERVLRSNMIDILPIL 240
Db 181 KVFLOGYQLHTLSJCNCSKSEEMITOLDQOVSSELCGLPKKRLAAERVLRSNMIDILPIL 240
Oy 241 RTLSTSPFPEKELAEATKTLHSGLTLAGLQELSMKRSMSMRQEWMLTVNNSSSSTOI 300
Db 241 RTLSTSPFPEKELAEATKTLHSGLTLAGLQELSMKRSMSMRQEWMLTVNNSSSSTOI 300
Oy 241 TKNLSTSHLPQHLAEATVTLDSLGLAQLSTKMSMSMRQEWMLTVNNSSSSTOI 300
Db 241 TKNLSTSHLPQHLAEATVTLDSLGLAQLSTKMSMSMRQEWMLTVNNSSSSTOI 300
Oy 301 YQAVSRIVCGHPEGGLKIKSLNMYEDNNYKALFEGNGTEDEATPYDNSTTPYCDNLK 360
Db 301 YQAVSRIVCGHPEGGLKIKSLNMYEDNNYKALFEGNGTEDEATPYDNSTTPYCDNLK 360
Oy 301 YQAVSRIVCGHPEGGLKIKSLNMYEDNNYKALFEGNGTEDEATPYDNSTTPYCDNLK 360
Db 301 YQAVSRIVCGHPEGGLKIKSLNMYEDNNYKALFEGNGTEDEATPYDNSTTPYCDNLK 360
Oy 361 NLESSPLSRITWKALKPLLVGKILYPTPATRYVAEYVKKTFQELAVFHDLEGMEELS 420
Db 361 NLESSPLSRITWKALKPLLVGKILYPTPATRYVAEYVKKTFQELAVFHDLEGMEELS 420
Oy 361 NLESSPLSRITWKALKPLLVGKILYPTPATRYVAEYVKKTFQELAVFHDLEGMEELS 420
Db 361 NLESSPLSRITWKALKPLLVGKILYPTPATRYVAEYVKKTFQELAVFHDLEGMEELS 420
Oy 421 PKITPMENSGEMLVNMLDSRNDHFWRQQLDGLDWTADQDIYAFIAKRPEDVQSSNGS 480
Db 421 PKITPMENSGEMLVNMLDSRNDHFWRQQLDGLDWTADQDIYAFIAKRPEDVQSSNGS 480
Oy 421 PKITPMENSGEMLVNMLDSRNDHFWRQQLDGLDWTADQDIYAFIAKRPEDVQSSNGS 480
Db 421 PKITPMENSGEMLVNMLDSRNDHFWRQQLDGLDWTADQDIYAFIAKRPEDVQSSNGS 480

```



GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 7, 2003, 08:50:05 ; Search time 48 Seconds
(without alignments)
1986.391 Million cell updates/sec

Title: US-09-595-526c-2
Perfect score: 11797
Sequence: 1 MACWPQRLRLMKNTLRRR.....VDAVLTSLQDEKESYV 2261

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 188354 segs, 42170167 residues

Total number of hits satisfying chosen parameters: 188354

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PC1_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PC105_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	11767	99.7	2261	10	US-09-995-542-11 Sequence 11, Appl
2	11767	99.7	2261	10	US-09-846-456-11 Sequence 11, Appl
3	10922	92.6	2201	10	US-09-995-542-9 Sequence 9, Appl
4	5888	49.9	2310	10	US-09-995-542-10 Sequence 10, Appl
5	5854.5	49.6	2273	10	US-09-995-542-12 Sequence 12, Appl
6	5768.5	48.9	2146	10	US-09-995-542-5 Sequence 5, Appl
7	5755.5	48.8	2144	10	US-09-858-194-2 Sequence 2, Appl
8	5731	48.6	2167	10	US-09-995-542-2 Sequence 2, Appl
9	5610.5	47.3	2121	10	US-09-995-542-6 Sequence 6, Appl
10	5581	47.3	2121	10	US-09-995-542-3 Sequence 3, Appl
11	4240.5	35.9	2436	9	US-10-156-239-8 Sequence 8, Appl
12	4240.5	35.9	2436	10	US-09-795-621-8 Sequence 8, Appl
13	4030	34.2	2001	9	US-10-072-621-8 Sequence 8, Appl
14	3832	32.5	1550	10	US-09-995-542-8 Sequence 8, Appl
15	2054	17.4	664	10	US-09-767-870-9 Sequence 9, Appl
16	1650	14.0	380	9	US-09-802-640-28 Sequence 2, Appl
17	1478.5	12.5	1642	12	US-10-090-458-5 Sequence 5, Appl
18	1470.5	12.5	1642	10	US-09-971-121-2 Sequence 2, Appl
19	1458.5	12.4	1638	12	US-10-090-458-2 Sequence 2, Appl

20	1390	11.8	1594	10	US-09-971-121-4 Sequence 4, Appl
21	1369	11.6	1624	12	US-10-090-454-2 Sequence 2, Appl
22	1368	11.6	1617	9	US-10-090-453A-2 Sequence 2, Appl
23	1311	11.1	1205	9	US-10-173-123-3 Sequence 3, Appl
24	1311	11.1	1207	9	US-10-173-123-3 Sequence 3, Appl
25	754	6.4	162	9	US-09-924-340-102 Sequence 10, Appl
26	754	6.4	162	9	US-09-924-340-102 Sequence 10, Appl
27	651.5	5.5	285	10	US-09-767-870-17 Sequence 17, Appl
28	518.5	4.4	1369	10	US-09-767-870-10 Sequence 10, Appl
29	460.5	3.9	199	10	US-09-767-870-18 Sequence 18, Appl
30	439	3.7	310	9	US-09-764-884-24 Sequence 24, Appl
31	439	3.7	310	9	US-09-860-670-93 Sequence 93, Appl
32	418	3.5	185	10	US-09-925-302-660 Sequence 660, Appl
33	358.5	3.0	340	9	US-09-738-626-5233 Sequence 5233, Appl
34	358	3.0	306	9	US-09-738-626-4554 Sequence 4554, Appl
35	340	2.9	292	9	US-09-738-626-5754 Sequence 5754, Appl
36	339.5	2.9	1280	9	US-10-072-621-7 Sequence 7, Appl
37	339.5	2.9	1280	10	US-09-866-866A-2 Sequence 2, Appl
38	339.5	2.9	1280	10	US-09-866-866A-4 Sequence 4, Appl
39	334	2.8	1272	10	US-09-769-097-4 Sequence 4, Appl
40	334	2.8	1272	10	US-09-815-242-10102 Sequence 10102, A
41	329.5	2.8	583	10	US-09-741-669-434 Sequence 434, Appl
42	327	2.8	894	10	US-09-815-242-10780 Sequence 10780, A
43	324.5	2.8	397	10	US-09-815-242-15842 Sequence 12842, A
44	324.5	2.8	423	10	US-09-815-242-15842 Sequence 14002, A
45	324.5	2.8	583	10	US-09-815-242-14002 Sequence 14002, A

ALIGNMENTS

RESULT 1
US-09-995-542-11
Sequence 11, Application US/09995542
Patent No. US20020127647A1
GENERAL INFORMATION:
APPLICANT: Shutter, John
TITLE OF INVENTION: ATP-Binding Cassette Transporter-Like Molecules and
TITLE OF INVENTION: Uses thereof
FILE REFERENCE: 00-658-A
CURRENT APPLICATION NUMBER: US/09/995,542
CURRENT FILING DATE: 2001-11-28
PRIOR APPLICATION NUMBER: 60/253,520
PRIOR FILING DATE: 2000-11-28
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 11
LENGTH: 2261
TYPE: PRT
ORGANISM: Homo sapiens
US-09-995-542-11

Query Match 99.7% Score 11767; DB 10; Length 2261;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 2254; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
QY 1 MACWPQRLRLMKNTLRRROTCOLLEAVAPLFIILISVRLSPYEQHECHPNKA 60
DB 1 MACWPQRLRLMKNTLRRROTCOLLEAVAPLFIILISVRLSPYEQHECHPNKA 60
QY 61 MPSAGTLPWVGIIICANNPCFRYPFGGAPGVGNFNKSIIVARLSDARLLYSOKPT 120
DB 61 MPSAGTLPWVGIIICANNPCFRYPFGGAPGVGNFNKSIIVARLSDARLLYSOKPT 120
QY 121 SMKMRVRLTLOQIKSSSNLKLQDFLVNEMFSGFLYHNLSLPKSTYDKMLRADVILH 180
DB 121 SMKMRVRLTLOQIKSSSNLKLQDFLVNEMFSGFLYHNLSLPKSTYDKMLRADVILH 180
QY 181 KVFLOGYQLHLTSLCNSKSEEMIQGDQVSESLCGLPKPKLAARVLRSMNDILKPTL 240
DB 181 KVFLOGYQLHLTSLCNSKSEEMIQGDQVSESLCGLPKPKLAARVLRSMNDILKPTL 240

```

OY 241 RLINSTSPSPSKELAEATKTLHSLGTLAQLFESKRSWSDKROEMFLTNVSSSSSTOI 300
DB 241 RLINSTSPSPSKELAEATKTLHSLGTLAQLFESKRSWSDKROEMFLTNVSSSSSTOI 300
OY 301 YOASRIVCGHPEGGLKIKSLNWEEDNNYKALFEGNGSTBEDAETFYDNSTTPYCNLDMK 360
DB 301 YOASRIVCGHPEGGLKIKSLNWEEDNNYKALFEGNGSTBEDAETFYDNSTTPYCNLDMK 360
OY 361 NESSPRLRIIMKALKPLLVKILYTPPTPATROYMAEVNTPQOLAVFHLEDMWELIS 420
DB 361 NESSPRLRIIMKALKPLLVKILYTPPTPATROYMAEVNTPQOLAVFHLEDMWELIS 420
OY 421 PRITFMENSOEMDLVRLMLDSRDNDHEWEOQLDMDTAODIYAFLAKHEDVQSSNGS 480
DB 421 PRITFMENSOEMDLVRLMLDSRDNDHEWEOQLDMDTAODIYAFLAKHEDVQSSNGS 480
OY 481 VYTMREAFENMOALRTISREMECNLNKLEPIATEVWLJNKSMLLDERKFMAGIYFTG 540
DB 481 VYTMREAFENMOALRTISREMECNLNKLEPIATEVWLJNKSMLLDERKFMAGIYFTG 540
OY 541 ITPGSELPHHYKIKRMDIDNVERTNKIKDGYMDPGRADPEEDMRVWGCPAYLDDVY 600
DB 541 ITPGSELPHHYKIKRMDIDNVERTNKIKDGYMDPGRADPEEDMRVWGCPAYLDDVY 600
OY 601 EOAIRVLTLGTEKKTGYVMQMPYCYVDITFLRVMSRSMPLMPLANTYSAVAVIKGIV 660
DB 601 EOAIRVLTLGTEKKTGYVMQMPYCYVDITFLRVMSRSMPLMPLANTYSAVAVIKGIV 660
OY 661 YKKEARLEETRMIMGLDINSILMFSEISLPIPLVSAGLVYLILKGNLLPYSDPSVYEV 720
DB 661 YKKEARLEETRMIMGLDINSILMFSEISLPIPLVSAGLVYLILKGNLLPYSDPSVYEV 720
OY 721 FLVSFAVYVTLLOCFILSTLFSRANLAACGGIYFTLYPLVLCVAMODYVGTLLKTFAS 780
DB 721 FLVSFAVYVTLLOCFILSTLFSRANLAACGGIYFTLYPLVLCVAMODYVGTLLKTFAS 780
OY 781 LLSPAFGGCEYFALFEEGOGIYQWMDLFEPSVEDEGNLTSTYSMMLPDTPLXGVMTW 840
DB 781 LLSPAFGGCEYFALFEEGOGIYQWMDLFEPSVEDEGNLTSTYSMMLPDTPLXGVMTW 840
OY 841 YLEAVFPGGYILPRMYPFCTKSYMFGESEDEKSHPSGNOKRISLICEEPTHLKLGVS 900
DB 841 YLEAVFPGGYILPRMYPFCTKSYMFGESEDEKSHPSGNOKRISLICEEPTHLKLGVS 900
OY 901 IONLYKVRDGMKAVANDGIALNFYEGOTTSLFGHNGAKTTMTSLTGLPPTSTATIL 960
DB 901 IONLYKVRDGMKAVANDGIALNFYEGOTTSLFGHNGAKTTMTSLTGLPPTSTATIL 960
OY 961 GKDISEMSTIRQNLGVCPOHNVLFDMLTVEEHIMFYARLKGISEKHAKEMQALDVG 1020
DB 961 GKDISEMSTIRQNLGVCPOHNVLFDMLTVEEHIMFYARLKGISEKHAKEMQALDVG 1020
OY 1021 LPSKLSKTSOLSGMOKLSVALAFVGSKVILDEPTAGVDPSRRGIWELLKTRQ 1080
DB 1021 LPSKLSKTSOLSGMOKLSVALAFVGSKVILDEPTAGVDPSRRGIWELLKTRQ 1080
OY 1081 GRTIILSTHMHMEADVLGRIALISHGKLCVGSGLFLKNOLGTGYTLTLVKKDESSL 1140
DB 1081 GRTIILSTHMHMEADVLGRIALISHGKLCVGSGLFLKNOLGTGYTLTLVKKDESSL 1140
OY 1141 SCRNSSSVYLLKEDSVQSSSDAGLSDHESDPLTIDVSAISMLIKKHVSEALVEDI 1200
DB 1141 SCRNSSSVYLLKEDSVQSSSDAGLSDHESDPLTIDVSAISMLIKKHVSEALVEDI 1200
OY 1201 GHELTYYVLPYBAKRGAYELFHEITDRLSDLGISSYGISETTLEIFLKAEEBGVDAB 1260
DB 1201 GHELTYYVLPYBAKRGAYELFHEITDRLSDLGISSYGISETTLEIFLKAEEBGVDAB 1260
OY 1261 TSDGTLPARNRARAGDDKOSCLRPTEEDAAPNDSDIDPESREFDLISGMDKGSYOVK 1320
DB 1261 TSDGTLPARNRARAGDDKOSCLRPTEEDAAPNDSDIDPESREFDLISGMDKGSYOVK 1320
OY 1321 GWKLTQOQFVALLMKRLILARRSRKGFPAQIVLPAVFCIALVFSLIIVPGKYPSELEQ 1380

```

```

DB 1321 GWKLTQOQFVALLMKRLILARRSRKGFPAQIVLPAVFCIALVFSLIIVPGKYPSELEQ 1380
OY 1381 PMWNEOQYTFVNSDADPEIDGTELLNALTKDPCFGRCMEGNPIPDTPQOAGEEEMTAP 1440
DB 1381 PMWNEOQYTFVNSDADPEIDGTELLNALTKDPCFGRCMEGNPIPDTPQOAGEEEMTAP 1440
OY 1441 VPOTIMDLFQNGNWMQONSPACOCSSDKIKKMLPVCPRGAGGLPPPOKQNTADILDL 1500
DB 1441 VPOTIMDLFQNGNWMQONSPACOCSSDKIKKMLPVCPRGAGGLPPPOKQNTADILDL 1500
OY 1501 TGRNISDYLVKYVQIIAASLKNK IWNNEFRIGGSLVSNTOALPPSOEVNDAIKOMK 1560
DB 1501 TGRNISDYLVKYVQIIAASLKNK IWNNEFRIGGSLVSNTOALPPSOEVNDAIKOMK 1560
OY 1561 HLKLAQSSADRLNSLGRMTGLDTRNNVKKWENNKGHAISFLNNALILRANLOK 1620
DB 1561 HLKLAQSSADRLNSLGRMTGLDTRNNVKKWENNKGHAISFLNNALILRANLOK 1620
OY 1621 GENPSHYGTTAFNHPILNTKQOLSEVALMTSDVLSICVIFAMSFPASVFEVLEIOER 1680
DB 1621 GENPSHYGTTAFNHPILNTKQOLSEVALMTSDVLSICVIFAMSFPASVFEVLEIOER 1680
OY 1681 VSKAKHLQFISGKPYTYMNSFWMDCMYVVPATLVITITFCOQKSVSSSTNLPLYAL 1740
DB 1681 VSKAKHLQFISGKPYTYMNSFWMDCMYVVPATLVITITFCOQKSVSSSTNLPLYAL 1740
OY 1741 LLLLGWSTIPLMYPASFEVKIPSTAYVLTSLYNLFTIGINGSATFVLELFDNKLNNIN 1800
DB 1741 LLLLGWSTIPLMYPASFEVKIPSTAYVLTSLYNLFTIGINGSATFVLELFDNKLNNIN 1800
OY 1801 DILKSVELFHPHCLGRLIDYKNOAMADALBERGENRVSPLSMDLVGRULFMAVVG 1860
DB 1801 DILKSVELFHPHCLGRLIDYKNOAMADALBERGENRVSPLSMDLVGRULFMAVVG 1860
OY 1861 VVEFLITVLQYRFFTRPRVNAKLSPLNDEDEDVREBORILIDGGGONDILEIKELTKI 1920
DB 1861 VVEFLITVLQYRFFTRPRVNAKLSPLNDEDEDVREBORILIDGGGONDILEIKELTKI 1920
OY 1921 YRRRRKPAVDRIKIPGCEFGLLGVNAGKSTPKMLTGTVTYRGDAFLNRNLSIN 1980
DB 1921 YRRRRKPAVDRIKIPGCEFGLLGVNAGKSTPKMLTGTVTYRGDAFLNRNLSIN 1980
OY 1981 IHEYHOMGICPOPDATITELLTGREHVEFPALLRGVPEKEVGVMARLKLGLVKYGEK 2040
DB 1981 IHEYHOMGICPOPDATITELLTGREHVEFPALLRGVPEKEVGVMARLKLGLVKYGEK 2040
OY 2041 YAGNYSGKNRKLSTAMALIGCPVVELEDEPTGMDPKARFLMNCALSVEKGRSVLT 2100
DB 2041 YAGNYSGKNRKLSTAMALIGCPVVELEDEPTGMDPKARFLMNCALSVEKGRSVLT 2100
OY 2101 SHSMECEALCTMAIMVNGRFRCLSVOHLKRRFDGTYIVRAGSNBDLKPVDFFG 2160
DB 2101 SHSMECEALCTMAIMVNGRFRCLSVOHLKRRFDGTYIVRAGSNBDLKPVDFFG 2160
OY 2161 LAFPGSVLKEKHNNMLQYOLPSSLSIARFSLISOSKRLIHEDYSVQOTLLDQFVNF 2220
DB 2161 LAFPGSVLKEKHNNMLQYOLPSSLSIARFSLISOSKRLIHEDYSVQOTLLDQFVNF 2220
OY 2221 AKQOSDDHLKDLSLKRNQTVVDVAVLTSFLQDEKVESEYV 2261
DB 2221 AKQOSDDHLKDLSLKRNQTVVDVAVLTSFLQDEKVESEYV 2261

```

RESULT 2

```

US-09-846-456-11
; Sequence 11, Application US/09846456
; Patent No. US20020146792A1
; GENERAL INFORMATION:
; APPLICANT: Rosier, Marie
; APPLICANT: Prades, Catherine
; APPLICANT: Lemoine, Cendriline
; APPLICANT: Naudin, Laurent

```

APPLICANT: Deneffe, Patrice
APPLICANT: Duveger, Nicolas
APPLICANT: Brewer, Bryan
APPLICANT: Remaley, Alan
APPLICANT: Foto, Silvia
TITLE OF INVENTION: Regulatory Nucleic Acid for the ABC1 Gene, Molecules Modifying It
FILE REFERENCE: 3606,0505
CURRENT APPLICATION NUMBER: us/09/846,456
CURRENT FILING DATE: 2001-05-02
PRIOR APPLICATION NUMBER: us 60/201,280
PRIOR FILING DATE: 2000-05-02
NUMBER OF SEQ ID NOS: 20
SOFTWARE: Patentin version 3.0
SEQ ID NO 11
LENGTH: 2261
TYPE: PR
ORGANISM: Homo sapiens
US-09-846-456-11

Query Match 99.7%; Score 11767; DB 10; Length 2261;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 2254; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 MACMPQLRLMLKMLTERRROTQCLLEVAMPFLFILILISRLSTPYEQHECHPNKA 60
DB 1 MACMPQLRLMLKMLTERRROTQCLLEVAMPFLFILILISRLSTPYEQHECHPNKA 60
QY 61 MPAGTLPWVGITICNANNPCFRYPTEGAPGVYGNFKSTVARLFSDARRLLYSKDT 120
DB 61 MPAGTLPWVGITICNANNPCFRYPTEGAPGVYGNFKSTVARLFSDARRLLYSKDT 120
QY 121 SMKDMRVLTLOQIKRSSNMLQDELVDNENFSGFLYHNLSPKSTVVKMLRADYILH 180
DB 121 SMKDMRVLTLOQIKRSSNMLQDELVDNENFSGFLYHNLSPKSTVVKMLRADYILH 180
QY 121 SMKDMRVLTLOQIKRSSNMLQDELVDNENFSGFLYHNLSPKSTVVKMLRADYILH 180
DB 121 SMKDMRVLTLOQIKRSSNMLQDELVDNENFSGFLYHNLSPKSTVVKMLRADYILH 180
QY 181 KVFLOQVQLHLTSLCNSKSEEMIQLODOEYSELGCLPKKELAAEVLVRSNMDILKPL 240
DB 181 KVFLOQVQLHLTSLCNSKSEEMIQLODOEYSELGCLPKKELAAEVLVRSNMDILKPL 240
QY 241 RLTNSTSPSPKELAEAKTLHSLGTLOELFMSRSMQROVMFLTNVNSSSTOT 300
DB 241 RLTNSTSPSPKELAEAKTLHSLGTLOELFMSRSMQROVMFLTNVNSSSTOT 300
QY 301 YOAVSRIVCGHPEGGLKISLAMYEDNNYKALFGNGTEDEATFPYDNSTTPYCNLDM 360
DB 301 YOAVSRIVCGHPEGGLKISLAMYEDNNYKALFGNGTEDEATFPYDNSTTPYCNLDM 360
QY 361 NLESSPLSRIIWKALKPLVGLKILYTPDPATROVMAEVNKTFOELAVFHDLEGMWELS 420
DB 361 NLESSPLSRIIWKALKPLVGLKILYTPDPATROVMAEVNKTFOELAVFHDLEGMWELS 420
QY 421 PKITMPENSGEMDLVRLDSDRNDHFWEQDLGDMTODIYAFIAKHPEDVQSSNGS 480
DB 421 PKITMPENSGEMDLVRLDSDRNDHFWEQDLGDMTODIYAFIAKHPEDVQSSNGS 480
QY 481 VYTWREAFNETNOAIRTISRMECVNLKLEPIATEYWLINKSMELDERKFMAGIYFTG 540
DB 481 VYTWREAFNETNOAIRTISRMECVNLKLEPIATEYWLINKSMELDERKFMAGIYFTG 540
QY 541 ITTSPSIELPHVVKIKIMQIDNVERTNKKIDGYWDPGRADPEEDMRYVMGGEAYLODV 600
DB 541 ITTSPSIELPHVVKIKIMQIDNVERTNKKIDGYWDPGRADPEEDMRYVMGGEAYLODV 600
QY 601 EOALIRVLTGEEKTGYYMOOMPVCYVDLFLVMSRSMPLFMTLAMIYSVANTIGIV 660
DB 601 EOALIRVLTGEEKTGYYMOOMPVCYVDLFLVMSRSMPLFMTLAMIYSVANTIGIV 660
QY 661 YEKEARLKETRINGLSDNLSILFMSFISLPLVLSAGLLVILKGNLLPYSDPSVVF 720
DB 661 YEKEARLKETRINGLSDNLSILFMSFISLPLVLSAGLLVILKGNLLPYSDPSVVF 720
QY 721 FLVSFAVYVTLIQCLISTLSRANLAACGIIYFTLYLPVLCVAMQDYVGFPLKIFAS 780

DB 721 FLVSFAVYVTLIQCLISTLSRANLAACGIIYFTLYLPVLCVAMQDYVGFPLKIFAS 780
QY 781 LLSPVAFGCGCEYFALFEEOGIGVQMDNLFESPVEEDGNLTSSIMMLFDFELGVMTM 840
DB 781 LLSPVAFGCGCEYFALFEEOGIGVQMDNLFESPVEEDGNLTSSIMMLFDFELGVMTM 840
QY 841 YIEAVFPGGYGIPRPWYFPCTKSVMFGEESDEKSHPGSNOKRMEISECEEBETHLKVS 900
DB 841 YIEAVFPGGYGIPRPWYFPCTKSVMFGEESDEKSHPGSNOKRMEISECEEBETHLKVS 900
QY 901 IONLVAVYVDGKAVANDGLALNFEEQITSEIGHGACGTTTMSILTGIFPTSGTAYIL 960
DB 901 IONLVAVYVDGKAVANDGLALNFEEQITSEIGHGACGTTTMSILTGIFPTSGTAYIL 960
QY 961 GKDIREMSTIKONLVCQOHNVLFDMLTVEEHIWFYARLKLSKHKYAEKQALDVG 1020
DB 961 GKDIREMSTIKONLVCQOHNVLFDMLTVEEHIWFYARLKLSKHKYAEKQALDVG 1020
QY 1021 LPSSKLKSTKSOLSGMOKRLSVALAFVGSKVYILDEPTAGVDYPSRGIWELLKRYO 1080
DB 1021 LPSSKLKSTKSOLSGMOKRLSVALAFVGSKVYILDEPTAGVDYPSRGIWELLKRYO 1080
QY 1081 GRTIILSTHMHDEADVLGRIALISHGKLCVGSLSFLKNOLGTGYTLTKKDYESSLS 1140
DB 1081 GRTIILSTHMHDEADVLGRIALISHGKLCVGSLSFLKNOLGTGYTLTKKDYESSLS 1140
QY 1141 SCRNSSSTVYLYKKEEDSVQSSDAGLSDHESDPTIDIVASINLIRKHVEARLVEDI 1200
DB 1141 SCRNSSSTVYLYKKEEDSVQSSDAGLSDHESDPTIDIVASINLIRKHVEARLVEDI 1200
QY 1201 GHELTIVLYPEAKEGAFVLEFHEIDRLSDIGISSYGETTLEEIFLKAEEGVDAE 1260
DB 1201 GHELTIVLYPEAKEGAFVLEFHEIDRLSDIGISSYGETTLEEIFLKAEEGVDAE 1260
QY 1261 TSDGTLPARNRRAFGDKOSCLRPETEDADPNDSIDIPESRETDLLSGMDKGSYQK 1320
DB 1261 TSDGTLPARNRRAFGDKOSCLRPETEDADPNDSIDIPESRETDLLSGMDKGSYQK 1320
QY 1321 GMLTIOQOEFVALMKRLLLARSRKGFQOIVLPAVFCIALVFSILVPPCKYSLELO 1380
DB 1321 GMLTIOQOEFVALMKRLLLARSRKGFQOIVLPAVFCIALVFSILVPPCKYSLELO 1380
QY 1381 PMWNEQYTFVNSDAPEDEGTLELNLALTKDPGFCRCMEGNPIPDFCOAGEEEMTAP 1440
DB 1381 PMWNEQYTFVNSDAPEDEGTLELNLALTKDPGFCRCMEGNPIPDFCOAGEEEMTAP 1440
QY 1441 VPOTIMDLFONGNMTQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPOKONTADILQDL 1500
DB 1441 VPOTIMDLFONGNMTQNPSPACQCSSDKIKKMLPVCPPGAGGLPPPOKONTADILQDL 1500
QY 1501 TGRNISDYLVKTYVOITIAKSLNKKIMVNEFRYGGFSLGVSNTQALPQSEVDAIKOKK 1560
DB 1501 TGRNISDYLVKTYVOITIAKSLNKKIMVNEFRYGGFSLGVSNTQALPQSEVDAIKOKK 1560
QY 1561 HLKLAQSSADREFLNSLGRFMTGLDTRNNVKKWPFNNKGHAISSEFLVNNAILRANLOK 1620
DB 1561 HLKLAQSSADREFLNSLGRFMTGLDTRNNVKKWPFNNKGHAISSEFLVNNAILRANLOK 1620
QY 1621 GENPSHYGITAFENHPLNLTAKOOLSEVALMTTSVDLVSLCVTFPAMSPVASFVFLIOER 1680
DB 1621 GENPSHYGITAFENHPLNLTAKOOLSEVALMTTSVDLVSLCVTFPAMSPVASFVFLIOER 1680
QY 1681 VSKAKHLQFLISGVKPYTYWLSNFMVDMCMYVVPATLVIIIFICFOOKSYVSSNTLPLVAL 1740
DB 1681 VSKAKHLQFLISGVKPYTYWLSNFMVDMCMYVVPATLVIIIFICFOOKSYVSSNTLPLVAL 1740
QY 1741 LLLLYGMSITPLMYPASFVKIPSTAYVVLTVSNLFTIGINSVAFVLELFDNKLNNIN 1800
DB 1741 LLLLYGMSITPLMYPASFVKIPSTAYVVLTVSNLFTIGINSVAFVLELFDNKLNNIN 1800
QY 1801 DILKSVFLIFPHECLGRLIDWKNQAMADALERGENFVSPUSMDVGRNLPMAVBC 1860

Db 1801 DILKSVFLIFPHFLCGLRGLIDMVKNOAMADALEREGERNFVPSLWDLVGRNLFAMAVEG 1860
 QY 1861 VVEFLITVLIOYRFFIRRPVNAKLSPINDEDEDYRREORHILIDGGONDLLEFLKELTKI 1920
 Db 1861 VVEFLITVLIOYRFFIRRPVNAKLSPINDEDEDYRREORHILIDGGONDLLEFLKELTKI 1920
 QY 1921 YRRKRRPAVDRIICVGIPEGCEGGLGVNAGKSSFTKMLTGDITVTTRGDAFLANKSLISN 1980
 Db 1921 YRRKRRPAVDRIICVGIPEGCEGGLGVNAGKSSFTKMLTGDITVTTRGDAFLANKSLISN 1980
 QY 1981 IHEVQNNNGYCPQOPATIELLIGREHVEFPALLRGVPEKEVGKGEAMIRKLVKVGKEK 2040
 Db 1981 IHEVQNNNGYCPQOPATIELLIGREHVEFPALLRGVPEKEVGKGEAMIRKLVKVGKEK 2040
 QY 2041 YAGNSGNNKRLSTAMALIGSPVFLDEPTTGMDPKARFELNCAISVKEGRSVLT 2100
 Db 2041 YAGNSGNNKRLSTAMALIGSPVFLDEPTTGMDPKARFELNCAISVKEGRSVLT 2100
 QY 2101 SHSMECALCTIRMAIMVNGRFRCLGSVQHLKNRFGDGTIVVRLAGSNPDLKPVDFEG 2160
 Db 2101 SHSMECALCTIRMAIMVNGRFRCLGSVQHLKNRFGDGTIVVRLAGSNPDLKPVDFEG 2160
 QY 2161 LAFPSVLEKEKRNMLQYQLPSSLSLARIFSILSOSKRLHIEDYSQOTLQVFN 2220
 Db 2161 LAFPSVLEKEKRNMLQYQLPSSLSLARIFSILSOSKRLHIEDYSQOTLQVFN 2220
 QY 2221 AKDQSDDDLKDLSLHKQTVVDVAVLTSFLQDEKVKESYV 2261
 Db 2221 AKDQSDDDLKDLSLHKQTVVDVAVLTSFLQDEKVKESYV 2261

RESULT 3

US-09-995-542-9
 ; Sequence 9, Application US/09995542
 ; Patent No. US20020127647A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Shutter, John
 ; APPLICANT: Ullas, laarni
 ; TITLE OF INVENTION: ATP-binding Cassette Transporter-Like Molecules and
 ; TITLE OF INVENTION: Uses thereof
 ; FILE REFERENCE: 00-658-A
 ; CURRENT APPLICATION NUMBER: US/09/995, 542
 ; CURRENT FILING DATE: 2001-11-28
 ; PRIOR APPLICATION NUMBER: 60/253, 520
 ; PRIOR FILING DATE: 2000-11-28
 ; NUMBER OF SEQ ID NOS: 24
 ; SOFTWARE: Patentin Ver. 2.0
 ; SEQ ID NO 9
 ; LENGTH: 2201
 ; TYPE: PRT
 ; ORGANISM: Mus musculus
 ; FEATURE:
 ; NAME/KEY: UNSURE
 ; LOCATION: (115)
 ; OTHER INFORMATION: amino acid at this position is unknown
 US-09-995-542-9

Query Match 92.6% Score 10922; DB 10; Length 2201;
 Best Local Similarity 94.9%; Pred. No. 0;
 Matches 2089; Conservative 55; Mismatches 57; Indels 0; Gaps 0;

QY 61 MPSAGTLFWOGIICNANPCFRYPPTGCEAPGVGNFNKSTIVARLFSDARLLYSOKDT 120
 Db 1 MPSAGTLFWOGIICNANPCFRYPPTGCEAPGVGNFNKSTIVARLFSDARLLYSOKDT 60
 QY 121 SMKDMKAVLRLOOTIKSSSNLKLQDFLVNERTSGFLIHNLSLPKSTVDKMLRADIYLH 180
 Db 121 SMKDMKAVLRLOOTIKSSSNLKLQDFLVNERTSGFLIHNLSLPKSTVDKMLRADIYLH 120
 QY 181 KVFLOGVYOLHTLSLNGSKSEEMIOLGDOVESEICGLPKREKLAARVLSNMOLILPIL 240
 Db 181 KVFLOGVYOLHTLSLNGSKSEEMIOLGDOVESEICGLPKREKLAARVLSNMOLILPIL 120
 QY 121 KVFLOGVYOLHTLSLNGSKSEEMIOLGDOVESEICGLPKREKLAARVLSNMOLILPIL 180

QY 241 RTINSTSPPEKELAEATKTLTLLSLCTLAOELFSMRSNMDROEVMFLTNVSSSSSTOI 300
 Db 181 TRKNSTSHLPQCHLAATVYLLDSLGIAOELFSKSNMDROEVMFLTNVSSSSSTOI 240
 QY 301 YQAVSRIVCGHPGGGLKIKSLNMYEDNNYKALFGNGTEDEAETFYDNSTPYCNLDLK 360
 Db 241 YQAVSRIVCGHPGGGLKIKSLNMYEDNNYKALFGNGTEDEAETFYDNSTPYCNLDLK 300
 QY 361 NLESSPISRIYKALPPLVYKILYTPDPTPATROVAEYKRFQELAVPHDEGMWELS 420
 Db 301 NLESSPISRIYKALPPLVYKILYTPDPTPATROVAEYKRFQELAVPHDEGMWELS 360
 QY 421 PKITWEMENSQEMDLVRMLDSRDNDHEWEOOLDGLMTADIVAFKAPREDVOSMGS 480
 Db 361 PKITWEMENSQEMDLVRMLDSRDNDHEWEOOLDGLMTADIVAFKAPREDVOSMGS 420
 QY 481 VYTMREAFNETNOAIFTTISRFMECVNLKLEPIPEVRLINKSMELDERKFMAGIVFG 540
 Db 421 VYTMREAFNETNOAIFTTISRFMECVNLKLEPIPEVRLINKSMELDERKFMAGIVFG 480
 QY 541 ITPGSELPHNVKKTIRMDIDNVERTNKIKDGYWDPGRADPFEDMRVYWGFAVLQDV 600
 Db 481 ITPGSELPHNVKKTIRMDIDNVERTNKIKDGYWDPGRADPFEDMRVYWGFAVLQDV 540
 QY 601 EOAIIIVLTGSEKKTIVYQOMPYCYVDIDILRWASRSMPLPMTLAMYSAVVIKGI 660
 Db 541 EOAIIIVLTGSEKKTIVYQOMPYCYVDIDILRWASRSMPLPMTLAMYSAVVIKGI 600
 QY 661 YEKEARKETMRIMGDINSILMFNFISSLIPLASAGLVIYILKGNLIPSDSVYFV 720
 Db 601 YEKEARKETMRIMGDINSILMFNFISSLIPLASAGLVIYILKGNLIPSDSVYFV 660
 QY 721 FLTSFAVNTILQCELLISTLESRANLAACGGITFTLYLVYLCVAMODVYGFPLKIRAS 780
 Db 661 FLTSFAVNTILQCELLISTLESRANLAACGGITFTLYLVYLCVAMODVYGFPLKIRAS 720
 QY 781 LSPVAFGCEFEALFEEOGIGVQMDNLFESPVEDGNNLTTSISMFLDFEFLGVMTW 840
 Db 721 LSPVAFGCEFEALFEEOGIGVQMDNLFESPVEDGNNLTTSISMFLDFEFLGVMTW 780
 QY 841 YIEAVEPGGYGIPRWPFPCTKSYWFEESDEKSHPGSNOKMSEICEEPEETHLKLVS 900
 Db 781 YIEAVEPGGYGIPRWPFPCTKSYWFEESDEKSHPGSNOKMSEICEEPEETHLKLVS 840
 QY 901 IONLVVYVYRDGMKVAVDGLALNFYEGQITSPFGHNGAGCTTMSILTGLPPTSGTAYIL 960
 Db 841 IONLVVYVYRDGMKVAVDGLALNFYEGQITSPFGHNGAGCTTMSILTGLPPTSGTAYIL 900
 QY 961 GKDIREMSTIRONLGVCPQHNVLPFMTLVEEHIMFYARKLSKRNHYAKEMQALDVG 1020
 Db 901 GKDIREMSTIRONLGVCPQHNVLPFMTLVEEHIMFYARKLSKRNHYAKEMQALDVG 960
 QY 1021 LPSSKLKSTISOLSGMOKKLSVALAFVGGSKVILLDEPTAGVDVPSRRGIWELLKTRQ 1080
 Db 961 LPSSKLKSTISOLSGMOKKLSVALAFVGGSKVILLDEPTAGVDVPSRRGIWELLKTRQ 1020
 QY 1081 GRTIILSTHMPADVYLGRIALISHGKLCYGVSSIFLKNQIGTGIVLTKKDDVSSLS 1140
 Db 1021 GRTIILSTHMPADVYLGRIALISHGKLCYGVSSIFLKNQIGTGIVLTKKDDVSSLS 1080
 QY 1141 SCNNSSTVSYLKKEDSVQSSSDAGLSDHESDPLTIDVSAISMLIRKHSEARLAVDI 1200
 Db 1081 SCNNSSTVSYLKKEDSVQSSSDAGLSDHESDPLTIDVSAISMLIRKHSEARLAVDI 1140
 QY 1201 GHELTIVLYPEAKKAGAFVFLHEIDRLSDGISSYGISETTLEIFLKAEEGVDAE 1260
 Db 1141 GHELTIVLYPEAKKAGAFVFLHEIDRLSDGISSYGISETTLEIFLKAEEGVDAE 1200
 QY 1261 TSDGTLPARNRRAFGDKQSLRPFTEDAADPNOSDDIPESRETDLLSGMDGKSYOVK 1320
 Db 1201 TSDGTLPARNRRAFGDKQSLRPFTEDAADPNOSDDIPESRETDLLSGMDGKSYOVK 1260
 QY 1321 GWKLTQOQFVALLMKRLLIARRSRKGFQAQIVLPAVFCIALVFSLIYPPFGKYPSELQ 1380

```

1261 GAKLTQOQFALLMKRLILARRSRKFFAOTVLPVAVFCALVFSLLVPCGYPSLEIQ 1320
1381 PMWNEQYTVSDADPEDTTLLELNALTQDPGFCRMCNGNIPDTPCOAGEENTTAP 1440
1321 PMWNEQYTVSDADPEDTTLLELNALTQDPGFCRMCNGNIPDTPCOAGEENTTAP 1380
1441 VPOTIMDLFONGWMTQNPSPACQSSDKIKKMLPVCPPGAGLPPOROKONTADILQDL 1500
1381 VPOSIVDLFONGWMTQNPSPACQSSDKIKKMLPVCPPGAGLPPOROKONTADILQDL 1440
1501 TGRNISDYLVKTYVQITIAKSLAKKIWNFRYGGFSLGVSNTQALPPSQEVDATKQMK 1560
1441 TGRNISDYLVKTYVQITIAKSLAKKIWNFRYGGFSLGVSNTQALPPSQEVDATKQMK 1440
1561 HLKLANDSSADREPLNSIGREMTGLDTRNNKYVFNKGMHAIISFLVNNATLRANLQK 1620
1501 LKLTMDTSADREPLNSIGREMTGLDTRNNKYVFNKGMHAIISFLVNNATLRANLQK 1560
1621 GENPSHYGTAFFNHPLNLTQKQLEVALMTTSVDLVSICVIFAMSPVAPSEVFLQER 1680
1561 GENPSHYGTAFFNHPLNLTQKQLEVALMTTSVDLVSICVIFAMSPVAPSEVFLQER 1560
1681 VSKAKHLQFTSGYKPYVITWLSNFMWMCNYPVATVITVITFCFOOKSVSSINLPLAL 1740
1621 VSKAKHLQFTSGYKPYVITWLSNFMWMCNYPVATVITVITFCFOOKSVSSINLPLAL 1680
1741 LLLLYGWSITPLMYPASFVKIPSTAYVVLTSYNLFIGINGSVATVLELFDNKNLNIN 1800
1681 LLLLYGWSITPLMYPASFVKIPSTAYVVLTSYNLFIGINGSVATVLELFDNKNLNIN 1740
1801 DILKSVPLTFPHFCGLGRGLDVMYKNQAMADALEFGENRFPSPULDVGRLFAMAVEG 1860
1741 DILKSVPLTFPHFCGLGRGLDVMYKNQAMADALEFGENRFPSPULDVGRLFAMAVEG 1800
1861 VVEFLITVLQYRFFTRPPRYNAKLSPLNDEDEDAVREKQRIIDGGGQNDILEIKELTKI 1920
1801 VVEFLITVLQYRFFTRPPRYNAKLSPLNDEDEDAVREKQRIIDGGGQNDILEIKELTKI 1860
1921 YRRRRKPAVDRIQVGIPECEFCGLGNGAGKSTFKMLTGPTTYVRGDAFLNKNSTLSN 1980
1861 YRRRRKPAVDRIQVGIPECEFCGLGNGAGKSTFKMLTGPTTYVRGDAFLNKNSTLSN 1920
1981 IHEVHONMGYCOFODAITELLTGREHVEFFALLRGVPEKEVKGEMATRKLGVLKYGK 2040
1921 IHEVHONMGYCOFODAITELLTGREHVEFFALLRGVPEKEVKGEMATRKLGVLKYGK 1980
2041 YAGNYSGGNKRKLTAMALIGGPVVEFLDEPTTGMDPKARRFLMNCALSVKEGKSVLT 2100
1981 YAGNYSGGNKRKLTAMALIGGPVVEFLDEPTTGMDPKARRFLMNCALSVKEGKSVLT 2040
2101 SHSMECEALCTRMATIMVNGRFRCLGSVOHLKNRFGDYTIIVIRLINGSNPDJLKVDPDFG 2160
2041 SHSMECEALCTRMATIMVNGRFRCLGSVOHLKNRFGDYTIIVIRLINGSNPDJLKVDPDFG 2100
2161 LAFPGSVLEKRRHMLQYOLPSSLSLARIIFSLSSOKRLHIEDYSVQOTLDOVFVNF 2220
2101 LAFPGSVLEKRRHMLQYOLPSSLSLARIIFSLSSOKRLHIEDYSVQOTLDOVFVNF 2160
2221 AKDOSDDHLKLDLHKNOTYVDVAVLTSFLQDEKKESEYV 2261
2161 AKDOSDDHLKLDLHKNOTYVDVAVLTSFLQDEKKESEYV 2201

```

```

RESULT 4
US-09-595-542-10
; Sequence 10, Application US/09995542
; Patent No. US20020127647A1
; GENERAL INFORMATION:
; APPLICANT: Shutter, John
; APPLICANT: Ullas, Jaarai
; TITLE OF INVENTION: ATP-Binding Cassette Transporter-Like Molecules and
; TITLE OF INVENTION: Uses Thereof

```

```

; FILE REFERENCE: 00-658-A
; CURRENT APPLICATION NUMBER: US/09/995,542
; CURRENT FILING DATE: 2001-11-28
; PRIOR APPLICATION NUMBER: 60/253,520
; PRIOR FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 2310
; TYPE: prt
; ORGANISM: Mus musculus
US-09-995-542-10

```

```

Query Match 49.9%; Score 5888; DB 10; Length 2310;
Best Local Similarity 50.1%; Pred. No. 0;
Matches 1168; Conservative 362; Mismatches 624; Indels 178; Gaps 27;

```

```

6 QLRLLMKNLFRRRQTCQLLEAVAMPILFILLISVRLSYPEYQHCHEPKNKAMPSAG 65
6 QLRLLMKNLFRRRQTCQLLEAVAMPILFILLISVRLSYPEYQHCHEPKNKAMPSAG 65
66 TLPMWQGIICNANNCFFRPTTPEAGPYVGNFNKSYARLFSDARLLLSQKDTSKDM 125
66 TLPMWQGIICNANNCFFRPTTPEAGPYVGNFNKSYARLFSDARLLLSQKDTSKDM 125
66 LLPWLQGIICNANNCFFRPTTPEAGPYVGNFNKSYARLFSDARLLLSQKDTSKDM 125
66 LLPWLQGIICNANNCFFRPTTPEAGPYVGNFNKSYARLFSDARLLLSQKDTSKDM 125
126 RVLERTLOOI-----KKSSNLKLODFLDNETPFCGLYHNLSPKSTVDKMLRAD 176
126 RVLERTLOOI-----KKSSNLKLODFLDNETPFCGLYHNLSPKSTVDKMLRAD 176
126 MAELRTLSQPMFDLRTHERFAGRLQIRDLKQDEBALTEFLMKRNTIGSDVAHLVNSQ 185
126 MAELRTLSQPMFDLRTHERFAGRLQIRDLKQDEBALTEFLMKRNTIGSDVAHLVNSQ 185
177 VILKRVPLQYQVQLHLSL-CNGSKSEMI---QIGQEVSE-LGGLPKKELAAEVR 230
177 VILKRVPLQYQVQLHLSL-CNGSKSEMI---QIGQEVSE-LGGLPKKELAAEVR 230
186 VVEFGVAGVDPDELDTIACSEALLQRFILFSQRRAGQVADALCPISQYTLQMIEDTLY 245
186 VVEFGVAGVDPDELDTIACSEALLQRFILFSQRRAGQVADALCPISQYTLQMIEDTLY 245
231 SMDLTKILTLNSTSPFPEKELAEATKTLHSLGTL--AQELFSKRW---SDMR- 282
231 SMDLTKILTLNSTSPFPEKELAEATKTLHSLGTL--AQELFSKRW---SDMR- 282
246 ADVDFK-----LFHVLPTLSSSGGIMLRMGCLISLSP 282
246 ADVDFK-----LFHVLPTLSSSGGIMLRMGCLISLSP 282
283 -----QEVNF-----LTNVSSSSSTQIYQAVSRIVCGPREGGLKIKSLMAYED 327
283 -----QEVNF-----LTNVSSSSSTQIYQAVSRIVCGPREGGLKIKSLMAYED 327
283 RMQKTIHPSVQDLWLRSPRLQNGGPEPTTQMLSTISDLGCTPBGGSFVSFNNYED 342
283 RMQKTIHPSVQDLWLRSPRLQNGGPEPTTQMLSTISDLGCTPBGGSFVSFNNYED 342
338 NNYKALFPGNGCTEBDAETFYDNTSTPYCNDLMKNLESSPLSRITWAKLPPLVGLIYTP 387
338 NNYKALFPGNGCTEBDAETFYDNTSTPYCNDLMKNLESSPLSRITWAKLPPLVGLIYTP 387
343 NNYKALFIDSTRKDPAYSDKRTTSPCNSLIQSLSPKTIAMRAKPLPMGLKILFTP 402
343 NNYKALFIDSTRKDPAYSDKRTTSPCNSLIQSLSPKTIAMRAKPLPMGLKILFTP 402
388 DTPAROVMAENVNTFOCLAFHDLGEMELSPKTYTPMNSQEMDLYMLDSDRNDH 447
388 DTPAROVMAENVNTFOCLAFHDLGEMELSPKTYTPMNSQEMDLYMLDSDRNDH 447
403 DSPAARRIKMKNANSTFEELDVKRLVKAWEVGPDIWFFEKSTQMTVIRDTLOHPYKD 462
403 DSPAARRIKMKNANSTFEELDVKRLVKAWEVGPDIWFFEKSTQMTVIRDTLOHPYKD 462
448 FWEQDLQDLMTADIVAFLLAKHEDVQSSNGSVYVTRAEAFNFTNOAIRTISPMCEVNL 507
448 FWEQDLQDLMTADIVAFLLAKHEDVQSSNGSVYVTRAEAFNFTNOAIRTISPMCEVNL 507
463 FTLNRLGSEGITTEAVLNFPNSNGPEQKQADMTSFDKRDIFNITDRRLANQYIECLV 522
463 FTLNRLGSEGITTEAVLNFPNSNGPEQKQADMTSFDKRDIFNITDRRLANQYIECLV 522
508 NKLEPIATEVWLINKSMELDERKFMAGIVFTGTPSIELPHVHKYKIRMDIDNVERTN 567
508 NKLEPIATEVWLINKSMELDERKFMAGIVFTGTPSIELPHVHKYKIRMDIDNVERTN 567
523 DKFESYDEVOLTORALSTLEENRPMAGVFPWMYPASSLPVHKYKIRMDIDNVERTN 582
523 DKFESYDEVOLTORALSTLEENRPMAGVFPWMYPASSLPVHKYKIRMDIDNVERTN 582
568 KIKQYMPGGRAPPEDEMRVWNGGFAYLDVYEQALIRVLITGEEKTGYVMOQMPPCY 627
568 KIKQYMPGGRAPPEDEMRVWNGGFAYLDVYEQALIRVLITGEEKTGYVMOQMPPCY 627
583 KIKQYMPGGRAPPEDEMRVWNGGFAYLDVYEQALIRVLITGEEKTGYVMOQMPPCY 642
583 KIKQYMPGGRAPPEDEMRVWNGGFAYLDVYEQALIRVLITGEEKTGYVMOQMPPCY 642
628 VDIIFLRVMSRMBPLFMTLWIVSVAVITIGIYVEKARLKEFTIRINGLINSILMF5MT 687
628 VDIIFLRVMSRMBPLFMTLWIVSVAVITIGIYVEKARLKEFTIRINGLINSILMF5MT 687
643 VDSFMIITLNRCPILPVLAMIVSVMTKIVGIVLEKELRLKELKNGVSNNAVIMCWFL 702
643 VDSFMIITLNRCPILPVLAMIVSVMTKIVGIVLEKELRLKELKNGVSNNAVIMCWFL 702
703 DSFMIITLNRCPILPVLAMIVSVMTKIVGIVLEKELRLKELKNGVSNNAVIMCWFL 762
703 DSFMIITLNRCPILPVLAMIVSVMTKIVGIVLEKELRLKELKNGVSNNAVIMCWFL 762
748 ACGGIYFTLVPLVLCVAMODVYGTLLKIFASLSLVAGFGCEYALFEEDGIGQMD 807
748 ACGGIYFTLVPLVLCVAMODVYGTLLKIFASLSLVAGFGCEYALFEEDGIGQMD 807
763 ACGGIYFTLVPLVLCVAMODVYGTLLKIFASLSLVAGFGCEYALFEEDGIGQMD 822
763 ACGGIYFTLVPLVLCVAMODVYGTLLKIFASLSLVAGFGCEYALFEEDGIGQMD 822

```



```

QY 808 NLESPVEEDGPNLTISIMMLEDFLYGVMTYEAEPGQYGIIPRWYPCOTKSYWFG 867
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 823 NICKSPLEGEDEFSLMSKMKMLLDALYGLLAWYLDQVPPGDIYPLPWYFLLOSWMYG 882
QY 868 -----EESDEKSHPSGNOKRMESEICOMEPEPHILKIGVSIOMLVKXY 908
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 883 GEGCSTREERALEKTEPLETEEMEDPEHP---EGMDSFFERELPGLVGVGVCKMLIKVP 938
QY 909 RDGMKAAVDGIALNPEYGOITSEFLGNAGAKTTTMSILGLEPPTSGTAYILGKDIRSEM 968
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 939 EPSGRRAVDRLNTFEYENQITAFELGNAGAKTTTSLILGLPPTSGVTLIGKDIETML 998
QY 969 STTRONLVCYPOHNVLPDMITVEEHTWETARLKLGSEKHYVKAEMOMALDVLCPSSKLKS 1028
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 999 DVVROSLGCMCPQHNIILFHHILTVAEHILFYAOLKGRSWEAEQLEMEMEDTGL-HHKRME 1057
QY 1029 KTSQLSGCMOKRLSVLAFVGGSKVYILDEPAGVDPYRGRIGWELLRYRGGRTIIST 1088
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1058 EADOLSGCMOKRLSVLAFVGGSKVYILDEPAGVDPYRGRIGWELLRYRGGRTIIST 1117
QY 1089 HHNDEADVLGDRIATISHGKLCVSSLEFLKNQDGTGYLLTVK--KDYES-----SLK 1140
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1118 HHNDEADVLGDRIATISHGKLCVSSLEFLKNQDGTGYLLTVK--KDYES-----SLK 1177
QY 1141 SCNNSSSTVYLYKKEDSVSSDAGLGDHSDTLITIVSALSILKHYSEARLYVEDI 1200
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1178 SCTSKGFTPCPTRYDEIT-----EEQVLDGVDVQELMDLVYHHVPEAKLVECI 1225
QY 1201 GHEITVLYPYEAKEGAFELFHEIDRLSDIGISSYGETTLEEFILKVAEESGVDAE 1260
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1226 GQELIFILPKNNKORAYASLFELEETLADIGLSSFGISDIPLEIFLKIYEDNAGAGSM 1285
QY 1261 TSGTGLPARNNRAFGDQSC-----LRPTE-----DDAADPDSOIDEPSRETDLL 1308
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1286 FVGG---AQOKRBOAGLRHPCASPTKRLQYADAPHTCSPGQVDPKQSPSEPED--- 1338
QY 1309 SCMDGKSYOVGMKLTQOQFVALLMKRLLIARSRKGFAGIIVDPAVVCALVLSLV 1368
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1339 -----PEVNTGARLLQHVQALLVKKRHHITRSKKDVAQVQVLPATVFLALMSTIV 1393
QY 1369 PPGKYPSLELOPMTNEOYTFVSNDAPEDTGTLELLMALITKDPGFGTROMECNPIPDTP 1428
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1394 PPGFEPALTLHPMYGHOYTFESMDEPNNEHLEVADVLNPPGCGNCLKEEMLEPEXP 1453
QY 1429 CQAGEEEMTAPYPTIMDFONGNMTMONPSPACOCSSDKIKMLPVCPPAGGILPPQ 1488
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1454 C-INATISWKTSPVSPNTTLHFQOKMTAAHPSPSCSTREKLTMLPECPREGAGILPPQ 1512
QY 1489 RKONTADILDLGRNISDVLVTVYVOIITAKSLKNKIWNEEFYGGFSLGVSTQALPPS 1548
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1513 RTQSTEVLODLNLRNISDVLVTVYPALIRSSLSKFWNEQYXGIGISIG-GKLPAPLIPS 1571
QY 1549 QEVNDALIKOKKHLKLAKOSSADRLNLSGR-----FMTGLDTRNN 1589
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1572 GE-----ALVGLSLGGMNNAVSGCVTRPASKEMLDPLKHELTTON 1613
QY 1590 VKVMENKKGHATSSPLNTVNNALIRANLOKGNPSHYGTANRHLNLTQOLSVALM 1649
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1614 IKWPFNNKGHALVSPFNVAHNAILTRASLPDRDPREYGTITVISOPLNLTKEDSLSTIV 1673
QY 1650 TTSVDLVLSICVTFAMSEVPASFVEIIOERYSAKHLQFISGVKPVIIWLSNFWDMCN 1709
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1674 TTSVDVAVALCVTFAMSEVPASFVEIIOERYSAKHLQFISGVSTYTWLNFMDIMN 1733
QY 1710 YVVPATLVITIFTCFOOKSVSTNLPVIALLLLYGWSITPLMTAFSEVFKIPSTAYV 1769
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1734 YAVSAGLVVIGFQOKKATSPDNLPALVSLMLYGMVAVIPMYAPASLFEVPSATYVA 1793
QY 1770 LTVSNLPTGINGSVAPFVLELPIDNK-LNNTINDILKSVFLPFPHCPGLIMYVNOQM 1828
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1794 LSCANLPTGINSATITVLELFEENKRLIRFNAMLRKLLIVFPHFLGKGLDLALISQV 1853
QY 1829 ADALERFGENRFVSPSLWDLVGRNLFAMAVEGVFLLITVLYOYRFFIRRPVNAKLSP 1888

```

```

Db 1854 TDVYAQGEESYANPFWMDILGKNLVMAIEGVVYPLTLTLLIHHFELRWIAEPAREPV 1913
QY 1889 NDEDEDVAREHORILIDGCGNDILEIKELTKIYRKRKRPADVRCIVGIPGECFGLGVN 1948
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1914 FDEDDVAAEERQVAMSGGNKTDILKLNELKIVYSGSSSPAYDLQVYRGCEFGILGVN 1973
QY 1949 GAGKSTTEKMLTGTITVTRGDALFNKNSILSNIHEVHONNGYCPQPDATITELLTGREHVE 2008
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1974 GAGTTFEKMLTGTITVSGDATVAGKSIILTSISDVQONNGYCPQPDADITELLTGREHVE 2033
QY 2009 FFALLRCYPEKEVKGEMARILKLVKYGKAYANGSGNKRRLSTALALTCGPLULL 2068
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2034 LYARIRGVPSKEIKVANWGIOSIGISLYADRLAGITVSGNKRRLSTALALTCGPLULL 2093
QY 2069 DEPTTGMDPKARFELNMCALSVYKESRVYLTSHSMECEALCTRMAIYNGRFRCLGSV 2128
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2094 DEPTTGMDPKARFELNMCALSVYKESRVYLTSHSMECEALCTRMAIYNGRFRCLGSV 2153
QY 2129 QHLKRRGDTYTVIVKLAGSN---POLKRYQDFPGLAFGVSILKEKRNMLDYOQLPSSL 2184
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2154 QHLKRRGDTYTVIVKLAGSN---POLKRYQDFPGLAFGVSILKEKRNMLDYOQLPSSL 2211
QY 2185 SSLARITSIISQSKRLHIEDYSQTTLDQVFNPAKQSDDDHLKDLISLH 2236
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2212 SSLARITSIISQSKRLHIEDYSQTTLDQVFNPAKQSDDDHLKDLISLH 2260

RESULT 5
US-09-995-542-12
: Sequence 12, Application US/09995542
: Patent No. US20020127647A1
: GENERAL INFORMATION:
:   APPLICANT: Shutter, John
:   TITLE OF INVENTION: ATP-Binding Cassette Transporter-Like Molecules and
:   TITLE OF INVENTION: Uses Thereof
:   FILE REFERENCE: 00-658-A
:   CURRENT FILING DATE: 2001-11-28
:   PRIOR APPLICATION NUMBER: 60/253,520
:   NUMBER OF SEQ ID NOS: 24
:   SOFTWARE: PatentIn Ver. 2.0
:   SEQ ID NO 12
:   LENGTH: 2273
:   TYPE: PRT
:   ORGANISM: Homo sapiens
US-09-995-542-12

Query Match          49.64: Score 5854.5; DB 10; Length 2273;
Best Local Similarity 49.94: Pred. NO. 0;
Matches 1163; Conservative 358; Mismatches 637; Indels 171; Gaps 27;

QY 6 QURLIMKNLTFRRQTCQLLEWAMPLEFLILISVRLSYRSPYEOHCFPNKAMPASG 65
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 6 QURLIMKNLTFRRQTCQLLEWAMPLEFLILISVRLSYRSPYEOHCFPNKAMPASG 65
QY 66 TLPPVQGITCANNDPCRYPPRPGAPGVGNPKNSIYARLEPSDARLLLYSOKDTSKDM 125
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 66 MLEPWLOGIFECVWNNPCFSPPGSPGIVSNYSILARYRDRDQELIMNAPESQHLGRI 135
QY 126 RKVLTJLQOI-----KSSSNLKLQDFLVNENFSGFLYNNLSLPKSTV---DKM 172
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 126 WTELHILTSQFMDTLRTIHERLAGGIRIKRDLKDEPILITFLINNIGLSDSVYVILLNSQ 185
QY 173 LRADVTLKRVFLQGYOHLTSL-CNGSKSEMIOLGQOEVE---LCGLPREKLAAB 226
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 186 VRPEGFANGV---PDLAKDIACEALIERFIISQRGBKATYRVALCSLSQGTQWIE 241
QY 227 RVLKRNMLILPILKTLNLSSTPFSKELAEATKTLULHSLGLADELKSMGSMW---SDM 281
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 242 DTLVANVDFK-----LFRVLPFLILDS---RSQGINLRSMGGLISDMS 281

```

Qy	282	-----ROEKMLFTNV----	-----NSSSTQIYQAVSNITGVGHGEGGKIKSLNWE	326
Db	282	PIQEFTHRPSMODLWTRPLMONGCPREFFTKIMJLSDLCGYPGEGSRVLSNWE	341	
Qy	327	DNNYKALFGNGTEDEAETFDNSTPTEPCNDLKNKNLESSPLSRITMKALKPLLVCILYT	386	
Db	342	DNNYKAFJGIDSTRKDPILYSDRRTTSFCNALIOSLESSNPLTFTIARRAKPPLMKCILYT	401	
Qy	387	PDTPATROYMAEVKTKTOELAVFHDLGEMBEELSPKITWTFMENSQEMDLVRLILDSRND	446	
Db	402	PDSPEARILIKNANSTEELERHAKLKAKEEVEGPOJIAWFEDNSTQMMNIRDTLGNPRYK	461	
Qy	447	HEWEOGLDGLMTADODIAFLAKHPEDYSSNCSYTAWTREAFNETQALRTTSIREFECVN	506	
Db	462	DELNUNQOGEESITAFDALINFLYKPRRSQADMDANPDWIRIFNITTRTLRLVNOYIECLV	521	
Qy	507	LNKLEPLATEWVLINKSMELDERKFMAGIVEMGITPGSELPHHVYKTRMDIDINVERT	566	
Db	522	LDKFESYNDETOLQORALSLLEENMAGVAFPDMMFWISLPRPHYKYKTRMDIDIVERT	581	
Qy	567	NKIKGXYNDPPRADPREDMDHMYGSGAYLODVEQAIITVLGTERTKGYVMQOMPYC	626	
Db	582	NKIKDRYDOSPRAADPVEDRFTIMGGAYLODMDEGITRSOYOAAPVGIYLOOMPYC	641	
Qy	627	YVDILFVMSRNPLEPFTLAMYISAVNIIKGLVYERKARLKTMRIMGLDINSILFMSF	686	
Db	642	FVDSEFMILNRCEPIFMVLAMYISVMTKSIVLEKELRLKTKLNGQSVNAVIMCTWF	701	
Qy	687	ISSLLPILVSGLLVYLKLCNLNLPYSDSEVNVFISVRAVVTIILOCELIITLFSRANLA	746	
Db	702	LDSFISMSMSELTTLTIFIMGRILNHSDDPLRPLFLAESTATIMCFLITPFEKASIA	761	
Qy	747	AACGGIYFTLYLRYEYVCAMODVGTETKIFASLSPAFVGGCGCFALPFEQOIGVOM	806	
Db	762	AACGSVITFTLYLPHILCFPMODRMAELKKAIVSLSPAFGCTGTLYVREFOGIGIOM	821	
Qy	807	DNLFESEVEDGFNLTTYSISMFLDFFLVGMWTYIEAVFPGOYGIIPREVPCTKSYWF	866	
Db	822	SNIGNSPTEGDEFSLSMQMMILDAVYDILLAMYDVOFPGCYGTPPLWYFILOESYWL	881	
Qy	867	G-----EESDEKSHPSNCKRMSELCIMEEPEHLKLGVSIONLKV	907	
Db	882	SGBGCSRTREBALEKTPLEETBEDPBP-----EGJHDSFEERHPGWGVYCNLVKI	937	
Qy	908	YRDGKVAVVDGLALNFEYEGOITSFLGHNGAGKTTTMSITLGTFPPTSGAYLYLGDIRSE	967	
Db	928	FEPGCRPAVRNLNTPFENQITAFGLGHNGAGKTTTILITGLLPTSGVLVYGRDIENS	997	
Qy	968	MSTIRONLGVCPOMNVTFLMLTYEENHWPYARLAKGISEKHVAEMQOMALDVGLEPSKUL	1022	
Db	998	LDAVROSIGKPOHNLTFHHLTYAEHMLFYAOLKGSQOEAOLEMEAMLEDTGL--HNRN	1056	
Qy	1028	SKTSQLSGQMOKRSLVALAVGVSGKVVILDEPAGVADPYRSRGIMELLYKTRQGTITLS	1087	
Db	1057	EEAODLSGQMOKRSLVALAVGVAKVYILDEPISGVADPYSRSSIMDLLLKYSGKTIIMS	1116	
Qy	1088	THHDEADVLDGRALITSHKLCVCVSSJFLKMQJDTGYTLTLVKKVDESSLSCSRNSSS	1147	
Db	1117	THHDEADVLDGRALITAGRIALICSGPTPLFLKPCGTGTLYTLVLR--MKNIOSORXSGEG	1175	
Qy	1148	TVSYLAKREDVSQSSDAGSGDSHESDTLIPVSAISNLIRKHSVAPRVEDIGHETLYV	1207	
Db	1176	TCSSSSKSGFS---TTCANHYDDTLPROVLDGYVNEIMDVYGLHHVPEAKKVCETIGELIFL	1232	
Qy	1208	LPIYAAKEGAFVELFHEIDRLSDLCITSSYGISSETTLLEEFLKVAEESGVDAETSDGTLP	1267	
Db	1233	LPINNFHRHRYASLFLRELETTIADLISFGISDTPLEEIFLKTVEDSDSGFLFAGG---	1289	
Qy	1268	ARRRRARFGDKOSLTPFTDDAANDNDSDI-----DPSESIEDLSGM	1311	
Db	1290	AQQKREYVNPRAHPLCP--REKACOTQODSNVCSFGAPAAHPEGQRPPEBECGDPJLNGT	1348	

Qy	1312	DQKSGYWKAKLHQOQFVLLMLKRLLIARSRKGFPAOIYIPAVVCAIAYFSIIVPPE	1371
Db	1349	-----OLVLOHVOALLKRRQHTIRSKHFLAOIVIPATFVVALMALSVILP	1397
Qy	1372	GKPYSELQPMWNEYOTFEVSNDAPEDTGTELELMAITKDPGTCRMEGNIIPDPCOA	1431
Db	1398	GEYPALTLHPVIYGOQYTFPESMDEPGEQYTVLADVLMLKPGGNMCKLEGMLEPYPC-G	1456
Qy	1432	GEEETTPAPVQOTIMDLFQNGNNTMONPSPACOSSDKIKKMLPVCPCGAGLPPRQKQ	1491
Db	1457	NSTPKRTPSVSPNITOLFQOKWOTVNPSPSCSTREKLTMLPECPBEGAGGLPPRQTO	1516
Qy	1492	NTADILQDTRNISDXLYLVYVOIILAKSLKNIMWNERYGGSLSYGNTQALPPSOEV	1551
Db	1517	RSTELQDLDTRNISDELVTYTPALLIRSLKSKFWNEORIGGISTG---GKLPVPIIT	1572
Qy	1552	NDAIKOMKRLHLAKDSSADRFNLISGR-----FMTGLDTRNNKV	1592
Db	1573	GEALV-----GLSLDGRIMNWSGGITPEAKSEIIPDLKHLIETEDNIKY	1617
Qy	1593	WPNKGMHAISSFLYNAINALIRANLOKGENSHGITAHPNHLPTKQOLSEVALMTTS	1652
Db	1618	WFNNKGMAHVLSEFLNVAHNAILRASLPKQDSPEYIGITVISOPLNTEQOLSEITVLPTS	1677
Qy	1653	VDVLVSCVIFPAMFVPSVSVFVLQDERSKAHNOFISGVAPVYIMLSNPFWMDCNIVY	1712
Db	1678	VDVAVALCVIFSMFVPSVSVLVLDERVKSXHOFGISGVPYTWYFNPLMDIMNVS	1737
Qy	1713	PATVILILFICFOOKSVYSSTNLPVALLLLLYGMSITPLAMPASFEKIPSTAVVLTSS	1772
Db	1738	SAGLIVGFLIGFOKKATTPSENIRPALVALLILYGMVITPMHPIASTFLPDVPSIAVALSC	1797
Qy	1773	VNLFITNGSVATFVLELFTDNK-LNNINDIKSVFLIPNHCILGRGLDIWYKQAMADA	1831
Db	1798	ANLEFGINSSAIFILELFPNNRTLLRFNAVNLRLKLIYEPHFCILGRGLDIALSQAVTDV	1857
Qy	1832	LEREGEKNRPSLMSDLYGNLIFAMAVEGVFELTVLQYRFFIRPRPVNAKSLPLNDE	1891
Db	1858	YARGEHEHSAHPFMDIDIGNLIFAMVEGVVFLTLVLQRFHFLISOWIAEPTKEPIYDE	1917
Qy	1892	DEDVREKORILDGGGNDLEIKELTKIYIRKRKPAVDRIQVIRPECFGLGAVNGAG	1951
Db	1918	DDVVAEBQRIRITGKNKTDILRLHELTIKIYLGSSPAVDRLCVGRPBGCGGLGAVNGAG	1977
Qy	1952	KSSFFKMLTGGDTTVTKGDAFLNKNSILSNIEHYHOMGVCPOFDALITELLTGREHVEFFA	2011
Db	1978	KTTTFKMLTGGDTTVTSGDAVNAKSKILTJNISSEHOMGVCPOFDALIDELTGREHLYXA	2037
Qy	2012	LLRGVPEKEVGKCEYAKIRGLIYKXGKVCYAGVSGGNKRKISTAMALIGGPRVYFLDER	2071
Db	2038	RLRGVPAEELIEKAVNNSIKSLGLITVYADCLAGYISGGNKRKISTAIALIGCPVLLED	2097
Qy	2072	TTGMDPKARREFLNMCALSVMKEGRSVLTSMSHEDEALCTPMALIMVNGRRCGLSGVOHL	2131
Db	2098	TTGMDPKARRLMNVVLSIIRGRAVVLTSMSHEDEALCTPLAIIIMVNGARCKMCTIOHL	2157
Qy	2132	KNRGQDGYTVIVKLAGSN---PDLPVODPFGLAEPGGSVLKEKHRLMLQYQOLBSSLSL	2187
Db	2158	KSKRGDGYIVTMKIKISFKDLDLPDLNPVEQFPQGNPGGVOREHRYNMLOFOVSS--SSL	2215
Qy	2188	ARIPISLISGKKRLHIEDYVSOTTLDOQVFNVPKQSDHDLKLSLH	2236
Db	2216	ARIPOLLISKSDLIEEYSVOTTLDOQVFNVPKQOOTES---HOLPH	2261

RESULT 6
 US-09-995-542-5
 Sequence 5, Application US/09995542
 Patent No. US20020127647A1
 GENERAL INFORMATION:
 APPLICANT: Shutter, John
 APPLICANT: Ullas, Iaarni
 TITLE OF INVENTION: AIP-Binding Cassette Transporter-Like Molecules and

QY 1907 GONDILEIKETIKYRRKRPVAVRIGVIGPECEFGLLGVNAGKSTFKMLTGDVTYV 1966
1788 TGGDVLTNLRNLYVGGQMPAVDRICIGIPGCEFGLLGVNAGKSTFKMLTGDVTYV 1847
QY 1967 RGDVFLNKSILSNHEVHQNNGYCPQPAITELLTGRHVFFALLRGVPEKEVKG 2026
1848 KGAVALAGHSVAREPSAALSMGCPQSDAIFELLTGRHELLRLRVRPAOVAQVAG 1907
QY 2027 MAIRKLGKYEKEYAGNYSNGNKRKLSTAMALLIGPPVFLDEPTTGMDFARRELVNC 2086
1908 SGARLGSLWYADRPAGITYSGGNKRKLATALLVGDPAVYFLDEPTTGMDFARRELVNC 1967
QY 2087 ALSVKEGSSVLTSHSMECEALCTRMALVNGRFRCLGSVQHLKKNRFGDYITVRIA 2146
1968 LLAUVREGSRVMTLSHMECEALCTRMALVNGRFRCLGSVQHLKKNRFGDYITVRIA 2027
QY 2147 GSNPDLKPYQDFPGALFPGSVYKEKRNMLQYOL-PSLSLARIFSLISQSKKLHIED 2205
2028 AAMS--QPAALVAAEPFESSELREAHGGRRLRQLPFGCRALAVFELAVHGAHEVED 2085
QY 2206 YSVSQTLQVEVNFAPKDDSDHKLKLSLHKNQTV-VDAV-----LTSFLQDEKVK 2257
2086 FVSQTMLEEVFLYFSKDGKDE---DTEQKEAGYGVDPALQHPKRVSOFLDDPSTA 2142
QY 2258 ES 2259
Db 2143 ET 2144

RESULT 7
US-09-858-194-2
Sequence 2, Application US/09858194
Patent No. US20020061590A1
GENERAL INFORMATION:
APPLICANT: GLUCKSMANN, MARIA
APPLICANT: CURTIS, ROBY A.J.
TITLE OF INVENTION: 38594, A NOVEL HUMAN TRANSPORTER AND USES THEREOF
FILE REFERENCE: MNI-153
CURRENT APPLICATION NUMBER: US/09/858,194
CURRENT FILING DATE: 2001-05-14
PRIOR APPLICATION NUMBER: 60/204,211
PRIOR FILING DATE: 2000-05-12
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 2144
TYPE: PR
ORGANISM: Homo sapiens
US-09-858-194-2

Query Match 48.8%; Score 5755.5; DB 10; Length 2144;
Best Local Similarity 50.2%; Pred. No. 0;
Matches 1146; Conservative 369; Mismatches 604; Indels 163; Gaps 27;

QY 301 YQAVSRIVCGHPEGGLIKISLNNYEDNNYKALFGNGTEDEAFETYNSTPYCDIIM 360
218 ---LSBALCSVRGSSSYVGPLNMYEASDLMELVG---QEPESALPDSSLSRASCSTLIG 270
QY 361 NLESSPLSRIIWKALKPLLVGKIYTPDPATQVMAEVNKTFOELAVFHDLGMMBEIS 420
271 ALDSHPILSRLLMRILKPLILGLKLFAPDPFTFRKLMQVNRTEFEELTLLDVREVEMLG 330
QY 421 PKITFMENSGEMLVNMLLDSRNDHFWEQDGLDWTMODIVAFILAKRPDEVOS---- 476
331 PRIFEMNDSSVAMLORLQLOMDEGR-RQPRGGR-----HMEALRSFLDP 377
QY 477 -SNSVYTRAFEMETQOAITISRFMECVNLKLEPIATEVWLINKSMELDERKFMAG 535
378 GSGG--YSWQAHADVGHVLTGLRTECLSDKLEAPSEALVSAIDLLAHEHRWAG 435
QY 536 IVFTGTPGSLTLPH-----HVKKYIRMDIDNVERTNKIKDGYWDPGRADPEEDMR 587
436 VVFLGPEDESSDPTDEHPITDLGPGHVIRIKIRMDIDVYTRTKIRDRFWDPGPADPLTDLR 495
QY 588 YWVGFAFALQDVNVOQAITIRVLTGTEKKTGYVMQMPYPCYVDOLFLRVMSRMLPFTTLA 647
496 YWVGFPYLLQDLVERAAVYLSGANPRAGLYLQMPYPCYVDVFLKRSLSPLFLTLA 555
QY 648 WYISVAVILKIGIVEKEARLKEETMRIMGLDINSILFESMFTISLPIPLVAGLVLILKLG 707
556 WYISVTLTVKAVAREKETRLDTRKAMGLSKAVYMLGMFLSCLGRLLSALLVLYLKLG 615
QY 708 NLDPYSDPSVVFVLSYPAVNTTILQCFILISTFESRANLAAACGGIITYFTLYPVILCVAN 767
616 DILYSHPGVVFLLFAAFVATVYQSFLLSAFSRANLAAACGGIITYFTLYPVILCVAN 675
QY 768 ODYVFTLIKIFASLISPARFGEYFALFEQIGVOMNLESPEDEGFMNLITSISM 827
676 RDRLPAGGRVAAVSLSPAFEGCESLALLEQEGQMHNVGTRP-ADVFSIAQVSG 734
QY 828 MLPTFLYGVMTWYIEAVFPQYGIIPRPWYFCTKSYWPFESDESKSHPSNORRMSEIC 887
735 LLDLAAVLGLATWYLEAVCPQYGIIPRPWYFCTKSYWPFESDESKSHPSNORRMSEIC 793
QY 888 MEEPTHLKLGVSITONLVKYYRDMGKVAVDGLALNTEGQITSLFGLHNGAGKTTMISLT 947
794 VEEAPRGLSPGVSVRSLEKRPSPOPALNGLSIDFYQHTIAFLGHSAGAKTTTISLS 853
QY 948 GLFPPSTGAYILKDIKRESESTIRONIGVOPHNVLFDMLTYEHIWFAIRLKGSEKH 1007
854 GLFPPSGSAFILLGHDKSSMAAIRPHLGYCPQYNVLFDMILYDEHWYFGRKLGLSAAY 913
QY 1008 VKAEMQWALDVGLPSSKRLKSKTSQLSQGMQRKLSVALAFVAGSKVYLIDEPAGVDPYS 1067
914 VGPEDDRILDQVGL-VSKQSVQVTRHLSGQMQRKLSVALAFVAGSKVYLIDEPAGVDPYS 972
QY 1068 RRGIMELLKTRQGRITILSTHNDVADVLGDRITAITSHGKLCCVSSLPFLKMLQGTGY 1127
973 RRGIMELLKTRQGRITILSTHNDVADVLGDRITAITSHGKLCCVSSLPFLKMLQGTGY 1032
QY 1128 LTLVKVDVSSLSGCRSSSVSVYLLKKEDSVSSQSAAGLSHDESPDTLITDVSAINLI 1187
1033 LTLVKARLPLTTN-----EKADTDMGSSVTRQEKKNGSGSAGVGPQOLLALV 1080
QY 1188 RKHVEARLVEDIEDHETLYVLYPEAKAGAFVLEHEDIRLSDLSISGISEETLLEI 1247
1081 QHWVPGARLVEELPHELVLVLYPYGADGSPATYFRRLDRLRLAELRLTGIGIDTSLLEI 1140
QY 1248 FLKYAESGVDVLETSDGTLPARRRNRAPG-DKQSCLPPTEDDAAPNDSDIDPESRET 1306
1141 FLKYAEECAADTDMEDSCGQHLCTGIAGLDVTLRMLMPQEPALF--NGEPAGSAPETD 1198
QY 1307 LLSGMDGKGYQVQVGMKLTQOQFVALLMKRLIARSRKGGPQIVLPAYFALVAFSL 1366
1199 QGSGPDVAVG--RVQGMALTRQOALILKRLRLARRSRGLFQIVLPALVAFVAFSL 1256

Db	213	PSSSIAPSCSELTGALDSHPRLSLRLMLRKLPLTLGKLLARPDPTRFKRLMAQVNRFEEL	27
Oy	407	AVFHDLECMWELSPKLTWTMENSQENDLVRLMLDSDNDHFMWQOOLDGLMTAADIAVAF	466
Db	273	TLLRRLREVMELPRLTFTFMNDSNVAALQRLLOMDDECR-RQPPGGRD-----	322
Oy	467	LAKHPEVDOS-----SNGSVYTWREAFENETNOALTRIISFMECVNLNKLEPIATEWMLIN	521
Db	323	---HMEALRSLFDPGSGG--YSMDADADVGHVGLTGHVTELESLDKLEAPSEALVLS	377
Oy	522	KSMELDERKMAAGIVETGIIPOGTELDH-----HKYKTRMDIDVERTNKIKOY	573
Db	378	RALQLEHNRMAQVYFLGDESDSPTEHPTRDLGEGHAIKIRMDIDVVTXNKIRDF	437
Oy	574	WDPPRADPFEDMTYWGGAVALODDYEAALITVLGTETKTCYVQWQYPCYVDITL	633
Db	438	WDPPRADPRLTDLRYWGGVYLODLVERAAVYLSGAPRAGLYIQQWRYCYVDVFL	497
Oy	634	RVMBSRMPLEMTLAMIYSVAVIIRKGIYEEKAREARLEKEMTRMGIDNSIMFWSFISLEPL	693
Db	498	RVLSSLEPLFTLLAMIYSVLTVAAVYREKETRLPTMRAMGSLRAYMLMFLSCLPF	557
Oy	694	LVSAGLLVILKLGKLNLPSPDSVYVFEIYSFVAVVITLCCFLISTFLFSRANLAAOAGGI	753
Db	558	LLSALLVLYLKLDILTPSHPGVWFLTAFAVAVTOSFELISAFSRANLAAOAGGLA	617
Oy	754	YFTLLYPLVLCVAMODVYGETLKIIFASLSPVAFGCEGFALFEEQOIGYVOMDLFESP	813
Db	618	YFSIYLPVLCVAMRDLRPLAGGRVAASLTSVPAGFGCESLALLDEQGGCAQWNVGTRP	677
Oy	814	VEEDGFNLITISIMKLDFTLYGKNTWYLEAVRPGQYGLPRMYFPCSTSYNPFGESEDK	873
Db	678	T-ADVFSIAQVSGLLLDAAVYGLATWYLEAVRPGQYGLPRMNFPRRSYSGPPRPS	736
Oy	874	SHPSQNRKMSLEIOMEEEPNLHKIAGVSIQWLVYVRDQMKVAVDGALAFNGOITSEFG	933
Db	737	PAPCTPLD-PRVLEYEARPOLSPGVSYSRLSRKFRPSPQRLRGSLDFYGHITAFLG	795
Oy	934	HNGAGKTTTMSILTLGFPPTSGTAYILGKDIRSEKMTIRONIGVCQOHNVLEFDMLTVEBH	993
Db	796	HNGAGKTTTSLISGLRPPSGSAFIIIGHVRSMSMAIRHILGVCQYUWVLEFDMLTVEBH	855
Oy	994	TWPAARLKGLSEKIVKEMEMDMLDVGRLPSKLTAKSTLSQSGGMOKRLVALLAPAGSKV	1055
Db	856	WVFGRKLKGLSAVYGEQDRLDQVGL-VKSOVOTRHLISGGMOKKLSVALAFVGGSGV	914
Oy	1054	VILDEPPAGVDPPYSRKGIIWELLKRYGRTIILSTHNMEDADVLGRIAIISHKGLCVG	1113
Db	915	VILDEPPAGVDPPARSRGIIWELLKYRGRTIILSTHNLDEALIGRVAVYVAGRGKLCGG	974
Oy	1114	SSLFLKNOIGIGYVLLVKKVDVSSSSGNSSTVSYLKKEDVSSOSSDAGLSDHSS	1177
Db	975	SPLELRRLSGYLLTVLTKARLPITTN-----EKATIDMGSGVDTRQEKKNS	1022
Oy	1174	DTLTIDVAISINLIRKHVSEARLEVEDIGHELTVLEYEAKEGAFVLEHEDIRLSLDG	1233
Db	1023	QGSRGFPQOLLATAYQHWVPARLVEELPHELTVLVTYTGADHSGFATLRFREDTRLELR	1082
Oy	1234	ISSYGISETTELEFLIKVAEESGVDAETSGCTIPARRNRRAG-DKQSLRPFTEDDAAD	1292
Db	1083	LTGTGIDSTLSLEELFLKVAEEDCAADDMEGSGGHLCTGIGLDTVTLRLKMPDETALF	1142
Oy	1293	PNSDDIPESREKTDILSGMGKSGYQVKGKTLQOQFVALLMKRLLIARSKRGFAVY	1352
Db	1143	--NBPAGSAPETDOGGSPRAYG--RYOGMALLRQOOLQRLILRLARSRGFLFADP	1198
Oy	1353	LPAVFCIALVFSILVPPFSKYPSLELQPMWINEQYTESNDAPEDTGTLELLNALTKDP	1412
Db	1199	LPALFVCIATVFSILVPPFGHYALRLSPMYGAQVSSFSADAPGRARLLEALLDEA	1258
Oy	1413	GFGRICMEGNIPDTPCOAGEEMTITAPVQGITMDLFQNGNNTMNOIPSPACCCSDKIKK	1472
Db	1259	G-----LEBPPVQHSNHRFSAPEYAEVAAKVLASGNMTPPSPSPACCCOSPGRAR	1308

QY	1473	MLPVCPCAGGLPRPOKKNOMTADLDDTGRNLSIDYLVTTYYOILIKSKTKNIWNEPRY	15
Db	1309	LLRPPCRAAAGGPRPPQAVVTSGGEVYONLIRNLSIDPLVTTYPRLYAQGLTKTKWNEARY	1366
QY	1533	GGESLGVSNTOALDPRSOEVNDATIKOMKKHLIKLAKDSSADRFNLISGKREMTGLDTRNNKY	1599
Db	1369	GGESLGVSNTOALDPRSOEVNDATIKOMKKHLIKLAKDSSADRFNLISGKREMTGLDTRNNKY	1422
QY	1593	WENKKGMAHLSFLNVTNNMLIRANLQKGNPSHYGTAFFNHPJLNTTKOOLESEVNLMTSS	1655
Db	1428	WENKKGMAHLSFLNVTNNMLIRANLQKGNPSHYGTAFFNHPJLNTTKOOLESEVNLMTSS	1487
QY	1653	VDVLVSTCVIFAMSFVASFVFLDIQERYSAKHIOFISGVKPRVLYWLSNPFYDMCNVY	1713
Db	1488	VDVLVSTCVIFAMSFVASFVFLDIQERYSAKHIOFISGVKPRVLYWLSNPFYDMCNVY	1544
QY	1713	PATVITITIFCOOKSVSYSTNNPVLALLLILGMSITPLMYPASVFKIPSTAVVLTSS	1772
Db	1548	PATVITITIFCOOKSVSYSTNNPVLALLLILGMSITPLMYPASVFKIPSTAVVLTSS	1607
QY	1773	VNEFIGINSVATFVLELFTDNKLNNINDILKSVLELFPHFCLGRGLDMVKNQAMADAL	1833
Db	1608	VNEFIGINSVATFVLELFTDNKLNNINDILKSVLELFPHFCLGRGLDMVKNQAMADAL	1667
QY	1833	ERPCENFVPSLSDILGRNLFPMANVGQVFLITVLIQYRFRIRRPVNAKLSPLNED	1892
Db	1668	ERLDDRFQSPFLREVYGVKKLLAMVLOGPLLELTLLLOHRSQOLLEPQVRSLLPLGEBD	1722
QY	1893	EDVRRERQRLIDGGGQNDILIEIKELTKYRKRRKPAVDRIQVIFPGECFGLIGVNGAGK	1952
Db	1728	EDVAREERERVQATQGDVLYLNRLLTKVYRGQRMPAVDRICLGIIPGECFGLIGVNGAGK	1787
QY	1953	SSTKMLTGGPTTYVIRGAFILNKSSILSNIEHVQNNXGQPOCDATTELLTGEHNEFPAL	2012
Db	1788	TSTFRMTYGDTLASRGVAIVLAGHSVAREPSAAHLSMGYCPQSDALFELLTGEHNEFPAL	1847
QY	2013	LRGVPEKVKGVKGEMALIRKGLGVKYEKYGAGNSGCKNRKRLSTAMALTGGPRVFLDEPT	2072
Db	1848	LRGVPEAQVQOTAGSSGLRATIGLSMADPRAGTYSGCKNRKRLATALLVGDPAVFLDEPT	1907
QY	2073	TGMPKARREFLNMALCSVYKEGSSVYLITSHMECCALCTRNAIMVNGFRCLSGVQHLK	2133
Db	1908	TGMPKARREFLNMALCSVYKEGSSVYLITSHMECCALCTRNAIMVNGFRCLSGVQHLK	1967
QY	2133	NREDDGTIVYIRIANGSNPDLKPVODEFFGLAFPGSVYLEKHKRNMLOYOL-PSLSLARIF	2191
Db	1968	GRFAAGHTLITLYRAAAMS-QPAAATVAIAEFPESELREAHNGRLRQOLPPGRCALAVIF	2022
QY	2192	SILSOSKRLIHIEDYVSQTTLDQVYVNFPAKXODSDDDHLKDLSLKNOTV-VDVAV----	2246
Db	2026	GELAVHGAHEGVDEPFSYSQMLEEVEFLYFSKDGKDE---DTEEOKEAGVGDPAPIQLOH	2082
QY	2247	--LTSFLODEKVKES 2259	
Db	2083	PKRVSOFLDDPSTAEIT 2098	
RESULT 10			
US-09-995-542-3			
Sequence 3, Application US/09995542			
Patent No. US20020127647A1			
GENERAL INFORMATION:			
APPLICANT: Shutter, John			
TITLE OF INVENTION: ATP-Binding Cassette Transporter-Like Molecules and			
FILE REFERENCE: Uses Thereof			
CURRENT APPLICATION NUMBER: US/09/995,542			
PRIORITY FILING DATE: 2001-11-28			
PRIORITY FILING DATE: 2000-11-28			
NUMBER OF SEQ ID NOS: 24			

QY 493 -----QAIRIS-----RMECVNLKLEPIATEWMLINKSME--LIDERKFWAG 535
 Db 548 LPSGMALLOQDLTDIDNACQIOFMKSVSDJFKGPFDEESTNYTLNAYQONVTFAS 607
 QY 536 IYFNGTLPSSIELPHYKXKIRMDINVERTNKIKDGYNDPGRADPFEDMKRYWNGCFAY 595
 Db 608 VITOTRKDGS--LPPHVKIKIRONSSTFEKTNIRAYWPGFNG--GREFYLGFWA 662
 QY 596 LODVDEOAILRVLGTE-KKTGVYMOOMPYPCYVDIPLRVMSRSPJLPMTLAMITYSAV 654
 Db 663 IODMERRAITDTPVGHVDEVPDGYQMPYPCTYTRDDLPJVEHMPILCMVTSWYSVAM 722
 QY 655 IIKGIYERKARKETMRIMGDNLSTLWFSWISLPLLVASAGLLVLTKGNLPSYD 714
 Db 723 TIOHIAEERHLEKVMKTMGINNAVHVAWFTTGVOLISVATLALTKGCVLISHH 782
 QY 715 PSVAVFELSVFAVVTIIOGFLISTLPSRANLAAAGGIIYFTLYLPYVLC-----VANOD 769
 Db 783 VYIMFLAVYAVATTIMCFPLSVLSKAKLASAGGIIYFTLYPVMYVAIREVAHDK 842
 QY 770 YVGTFLKIFASLLSPVAFGECEYFALFBEQIGVOMDLFESPYEEDGFNLTSISMIL 829
 Db 843 ITAFE-KCIASLMSSTAAGISKXFALYEVAGVGIQWHTFSSQSPYEGODFNLLAVTMM 901
 QY 830 FDFELGVMTWTYEAFFPOYGIIPRWYTPCTKSTWFE--ESDEKSHPGSNOKRMS-- 884
 Db 902 VDAVAVGILTWYEAHVHPMYGILPRPWPYPOKSTWLSGSRTEAWMSWPMARTRPLSYM 961
 QY 885 ---EIC-----NEEPTHLKLGVSIONLVKVRDGMKAVIDALANFEGOT 929
 Db 962 EEDQACAMESRREETRGMEEPETHLPVVCYDKLTKYKDKKLALINKSLINLEYNOV 1021
 QY 930 SEFGHAGAKTTTMSLJGLFPPTSGTAYILCKDIRSEKSTIRONLGYCPOHNVLEDMIT 989
 Db 1022 SEFGHAGAKTTTMSLJGLFPPTSGTAYILCKDIRSEKSTIRONLGYCPOHNVLEDMIT 1081
 QY 990 VEEHMFYARLKLSEKHYKAMEQMALDVGJPSKLSKTSQSLSGOMRKSTVALAVG 1049
 Db 1082 VEEHMFYARLKLSEKHYKAMEQMALDVGJPSKLSKTSQSLSGOMRKSTVALAVG 1140
 QY 1050 GSKVYVLDPTAGVDPYSRKIMWELLVYRGRTIILSTHMDADVLDGRIATISHGL 1109
 Db 1141 GSRATILDEPTAGVDPYARARIMDLILKYPGRITILSTHMDADVLDGRIATISHGL 1200
 QY 1110 CCVGSSEFLKNOLOGYVYLLVYKDVESLSCSRSSSTVSTLKEDSVSSSSAGJCS 1169
 Db 1201 KCCGSEPLFKGTGGYKRLTLVKKPAEPG-----GPOEPGLAS 1238
 QY 1170 DHESDTLTIDVS--AISNLIRKHSSEARLVEDIGHELTYVLYPEAKEGAFVLEFHEID 1227
 Db 1239 SPGRAPLSSCSELOVSGFIRKHVASCLLVSTSTELSYILPSEAKAGCAERLPHLER 1298
 QY 1228 RLSDLGSSYISSETTLEEFILKVAE-----SGVDA-ETSDGTLP----- 1267
 Db 1299 SUDALHSSFGIMLDTTLEEVFLKVSSEDOJLSESDADVKESRKDVLPGAEGPASGEGHAG 1358
 QY 1268 -----ARRNRR--FGDRQSLRPTEDDADPNDSIDPEPSR 1303
 Db 1359 NLARCELTQOSASTQASVSGASAGDEGAGTYDYGYRPLF--DNODDP--NVSJDEV 1415
 QY 1304 ETDLLSGMDGKGSYOVKWKLTQOQFVALLKRLILARSRRKGFPAOTVIRPAVEVCIJALV 1363
 Db 1416 EAEALSRV--GQSSRKLDGWLKVRQFHLLVKKRHCARNSKALFESQILLPAFVCAVMT 1474
 QY 1364 FSLVPPGKVPSTLELOPMWYNEQYT-----FVSDAPE-----DTGTELLNAL 1408
 Db 1475 VALSPETIGDLPVYLSQTH--NTTOPRGNFIPYANERREYRLRSLSPASSPOOLSTF 1533
 QY 1409 TKDPGFGTCM-----EGNPI----- 1424
 Db 1534 RLPSGVATCVLKSPANGSLGPTLNLSSGESRLAARFDSMCLESTFQIGLPLSNFVPPP 1593

QY 1425 -----PD-----TPQAGFEWETAP--VPOTIMDLFONGNMTQONPSPAC 1463
 Db 1594 PSPAPSDSPAPDDDLQAMWNSLPTTAGPEMKTISAPSLPRVREPRV-----C 1641
 QY 1464 QCSSDKIKKMLPVCPGACGILPPQOKONTADILQDLTGRNISDYLVKTYVOJIAASLKN 1523
 Db 1642 TCSAGTGFS--CPSSVG--HPQOMRVVTGDLITJDTGHVNSRYLLFTSDRF----- 1690
 QY 1524 KIWVEPFYGGFSLGVSNTQALPSSQEVNDAIKMKHKLKLANDSSADRNLSTIGREMTG 1583
 Db 1691 -----RLHRYGALITG--NLKSTPASTCTRAPPRVKA----- 1721
 QY 1584 LDTNRNKKVWENKNGMAHISSELVNINNALRANLOKE--NPSHYGITAENHPLNLTQO 1642
 Db 1722 IAVBRAQVFNKNGYSMPTLNSLNNALLRANLPRSKGNPAAYGTTVNHMMNTSAS 1781
 QY 1643 LSEVALMTSYDVLSVCVIFAMSFVASTVYFLIOERYSKAKHLOPISGKPVYIWLSN 1702
 Db 1782 LS-LDYLGOTDVVAIAIFIIVAMSFVASTVYFLVIAEKSTAKHLOPISGKPNIIYWMAN 1840
 QY 1703 FVMDNCVNVVPAITVITIFCFQOKSVSSNLPVALLLLLYGWSITPLMYPASVFKI 1762
 Db 1841 YVMDMLNVLPAITCVITILYFVFLDPATTPNPPAVSLFLYLGWSITPLMYPASVFKI 1900
 QY 1763 PSTAYVVLTSYNEFIGINGSVATFVLELT--DNKLNNINDILKSVPLIPPHFCGRGLID 1821
 Db 1901 PSSAYVFLVYINLFIATATVATFLLQLFEDHDKLVNSYLSKSCFLIPNVLHGGLME 1960
 QY 1822 MKRQAMADALERFE--NREVPISMDVLGRNLFAMAABEVYFLLTYLOFETFRPP 1880
 Db 1961 MAYETINETYAKIGQDPMKSPENDIVTRGLVMAABEVYFLLTYLOFETFRPP 2020
 QY 1881 VNAKSLPNDDEDVREERORILDGGGONDILEIKELTKIYRKK--RKPAVRIKVGIP 1937
 Db 2021 MPVSTKVEVD--DVDVASERRVLRGADNDNMVAKIENLTKYKSKRIKGRILLADRLCLAR 2079
 QY 1938 PGCEFGLLYNGAGKSTFKMLGDTTIVTRGDAFLKNSITLSNINEHONMGYCPQFDAI 1997
 Db 2080 PGCEFGLLYNGAGKSTFKMLGDTTIVTRGDAFLKNSITLSNINEHONMGYCPQFDAI 2139
 QY 1998 TELJTGREHEEFALLNGVEKEVKGAWATRLGLVYKGEYVAGNYSNGNKRKSTAM 2057
 Db 2140 FDELTAREHELOITRLRGLISMKDEARVYKALEKLTLTADKPACTYSGGNKRKSTAI 2199
 QY 2058 ALLGPPVYVLEPTTGMDBKARFLMNCALSVYKGRSVYLTSHSMECEALCTRMATM 2117
 Db 2200 ALLGPPVYVLEPTTGMDBKARFLMNCALSVYKGRSVYLTSHSMECEALCTRMATM 2259
 QY 2118 VNGRFGISVQHLKRNFGGYTIVRIRAGSNPDLPYODFGILAPPGSVLKEKRNMLQ 2177
 Db 2260 VNGRFGISVQHLKRNFGGYTIVRIRAGSNPDLPYODFGILAPPGSVLKEKRNMLQ 2318
 QY 2178 YOLPSSLSLARIETLSQSKRRLIEDYSVSQTLTDOQVFNFAKQSDQD 2227
 Db 2319 YOLKSEHISLAOVESKMEQVSGVLGIEDYSVSQTLTDOQVFNFAKQSDQD 2368

RESULT 13
 US-10-072-621-8
 ; Sequence 8, Application US/10072621
 ; Patent No. US20020169137A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Relinet, Peter B.
 ; APPLICANT: Comop, Bruce P.
 ; APPLICANT: Pollard, Michelle
 ; TITLE OF INVENTION: REGULATION OF AMYLOID PRECURSOR PROTEIN EXPRESSION
 ; FILE REFERENCE: 100103.402
 ; CURRENT APPLICATION NUMBER: US/10/072,621
 ; CURRENT FILING DATE: 2002-02-08
 ; NUMBER OF SEQ ID NOS: 10
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 8

: LENGTH: 2001
 : TYPE: PROT
 : ORGANISM: Homo sapiens
 : FEATURE:
 : NAME/KEY: VARIANT
 : LOCATION: 30, 70, 280, 477, 558, 1471, 1651, 1689, 1724
 : OTHER INFORMATION: xaa = Any Amino Acid
 : FEATURE:
 : NAME/KEY: VARIANT
 : LOCATION: 30, 70, 280, 477, 558, 1471, 1651, 1689, 1724
 : OTHER INFORMATION: xaa = Any Amino Acid
 : US-10-072-621-8

Query Match 34.28; Score 4030; DB 9; Length 2001;

Best Local Similarity 43.8%; Pred. No. 5, 2e-311;

Matches 886; Conservative 294; Mismatches 549; Indels 296; Gaps 46;

QY 382 KILTPPTATROYMAENVKTFQELAFVHLEGMWELSKITFMENSOEMLVRLLD 441
 DB 26 KILYXPGSEYDRVILKANETFAVGNVTHYAQVWLNISAEIRSKLEQGRLOHLRL-- 83
 QY 442 SRDNHFWEOQLDLDWTADQIVAFIAKHPEDVYSSNGSVYTWREAFNETN----- 492
 DB 84 -----QYVAELRPHPE---ALNLSTDELPPALRODNESLPSGML 121
 QY 493 -QAIRTIS-----RMECVNLNKLEPIATEVWLINKSME--LDERKFWAGIVFTGIT 542
 DB 122 LQOQDITDNPACGMIOFMSKVSDIFKGFPEDESIVNYTLNOAYQDNVTVFAGVIFQTRK 181
 QY 543 PGSLIEPHHAKYKIRMDIDVVERTNKIKDGYMDPRADPFEDMRKVMGFALDQVVD 602
 DB 182 DGS--LRPHHYKIRONSSTFTEKTEIRRAYWRPGNTG---GRFYLGFWIDOMMER 236
 QY 603 AIIIRVLTEG-KRTGYVYMOQMPYPCYVDLIFLRVMSRMPLEMTLMIYSAVVIKGIY 661
 DB 237 AIIIDTEFGHUVNPPGYSQVMPYPCYTRDPLFVIEHMMPLCXISWYVSVAFTIOHIVA 296
 QY 662 EKEARLEKEMRINGLONSTIMESWFTLSLPLVLSAGLLVLIKGNLLPYSDPSVVEVF 721
 DB 297 EKHHRLKEVAKTKGLNNAVHVAWAMFTTFVQLSISVATLAIKYGQVIMHSHVVIWLF 356
 QY 722 LSVFAVVTILQCLFSLTSLSRANILAAAGGIIYFTLYLPLVLC-----VAMQDYVFTLK 776
 DB 357 LAYVAVATIMFCFLSVLVISKAKLASA-GGIIFLSTVYMYVAITEEVAHDKITAFE-K 414
 QY 777 IFASLLSPAFAFGGCEYFALFEQIGVQVMDLFEPSVEEDGFNLITTSISMFLFTFLYG 836
 DB 415 CIASIMSTTAFGLGSKYFALYEAVAGVGIOMHTFSSQSPVEGDENLLAVTMLVDAVVYG 474
 QY 837 VMTWYTEAVEPPOGYIPRMYRPTCTSYWFG-----ESDEKSHPSQOKRMS-----ETC- 887
 DB 475 ILKMYTEAVHPKMYGLPRMYRFPLOKSYWLGSRTEAMWMSWMAWTPRLSVMEEDQACA 534
 QY 888 -----MBEPTHLKLGVSIONLVKYYRDGKRVADVGLALFYGGQITSLFNGH 936
 DB 535 MESRFEETRGMEDEEPTHLPLVYVYVXDKLKYKDDKKLANKLSLNLVYNOGVSEPLGNG 594
 QY 937 AGCTTMSLTLGLFPPTSGTAYILGKDISEMSTTRQNLGVCPOHNVLFDMLTVEHLMF 996
 DB 595 AGCTTMSLTLGLFPPTSGSATIYGHDIRTEDEIRKN-GHVPQHNVLFDRLTVEHLMF 653
 QY 997 YARLKLSEKHYKAEQMALDVLGSPSKLSTKTSOLSGMGKRKISVALAFVGGSKVYL 1056
 DB 654 YRLKSMQOEELPREMDKMTEDLEL-SNRHSLVOTLSGMRKRVSAIAFVGGSKRAILL 712
 QY 1057 DEPTAGVDVYSRGGIWEILLKRYROGRTIILSTHMDADVLDGLAIITSHGKLCVGSLL 1116
 DB 713 DEPTAGVDVYARRAILMDLITKYKPGRTILSTHMDADLGDRLAIITSHGKLCVGSPL 772
 QY 1117 ELKNQUGCYVYTLVKKVYESSLSGRSSSTVYTLKREDSVSSSAGLCSHESPTL 1176
 DB 773 FLKGYGDSYRLTLVKKRAEPG-----GPOEPGLASSPGRAP 810

QY 1177 TIDVS--AISNLIRKRVSEARLVEDIGHETLVLPYEAAGKGAPELFEHIDRLSDGI 1234
 DB 811 LSSCSLOVSOPIRKHVASCILVSTSTELSTLSESAKGAERLRLQHLERSIDLALH 870
 QY 1235 SSTYSETTLEETFLKVA-----EESGVA-ETSDGTLPARNR----- 1272
 DB 871 SSFGLMDTTLLEVFLKVSQDGLSNGADVYESKRDVLPAGEHNASGEGHAGNARCE 930
 QY 1273 -----RAGDKOSCLRPETE-----DDAADPNOSIDPESRETDL 1308
 DB 931 LTIQSASLOSASSVSGALDEGA---GYTDVYGDYPLPLDNPDPDP--NLSLOVEAEAL 985
 QY 1309 SGMDKGSYQVKGMLTDOQFVALLMKRLRLARSRKCFPOIYLPAYFCIALVFSILY 1368
 DB 986 SNV-CGGRKLDGMLKQFHLGFLKRFHARNSKALFOLLPAFFVCVAIVTALSV 1044
 QY 1369 PPGKYPSTLEQPMYNEQYT-----FVSNDAPE-----DTGTELLNALTRDPG 1413
 DB 1045 PEIGDLPLVLSPSQYH-NYTOPRGNFIPYANEEERERLRLSPDASPOOLVTFRLPSG 1103
 QY 1414 FGTRCM-----EGNPI----- 1424
 DB 1104 VQATCVLSPANGSLGPTLNLSSGESRLAARFDSMCLESFTGLPLSLNFVPPPSAP 1163
 QY 1425 -----PD-----TPQAGEEETTPAR-VPOTIMDLFONGMWMTQNSPACQSSD 1468
 DB 1164 SDSPASPDDELDQAMVSLPPTAGQEMTSAPSLRLVNEPR-----CTCSAQ 1211
 QY 1469 KIKKMLPVCPPGAGLPPPOKQONTADILQDITGRNISDYLVKTYVOITAKSLKNKIVN 1528
 DB 1212 GTGFS---CPMSVGG-HPPOMRVYTGDIITDIGHNVEYLLFTSDPR-----R 1256
 QY 1529 EFRYGFSLGVSNTQALPPRQSOEVNDALIKOMKHLKLAKDSSADRFNLSLGRFMTGLDTRN 1588
 DB 1257 LHRVATITFG-NVLKSTFAS--FGTRADPMYKTRCANA----- 1293
 QY 1589 NVKVMENKGMHAISSPLNVNINAILRANLOKGE-NESHYGFITAFNHPNLTKOOLSEVA 1647
 DB 1294 --QVYFNKKGYSMPYTLNSLNNAILRANLPSKGNPRAAGITTYTNPMKTSLSLSD 1350
 QY 1648 LMTTSVDVLSICVIFAMSEVPASVVFVFLQERYSKAKHLQFISGVKPVYIWLNSFVMD 1707
 DB 1351 YLQGTDVVAIFITIVAMSEVPASVVFVFLVAKSTKAKHLQFISGVKPVYIWLNSFVMD 1410
 QY 1708 CNYVPATLVIIIFICFOKSYVASTNLPLVALLLLIYGSITPLMYPASVFPVPTAY 1767
 DB 1411 LNYLVPATCCVILFVFDLPAYTPTNFPAVLSFLHGMSTPLMYPASVFPVPTAY 1470
 QY 1768 VVLSVNLFIGINGSVATFVLELT-DNKLNNINDILKSVELFPHFCGLGRLIDWVKNQ 1826
 DB 1471 XFLIVINLFIGITATVATFELLQFEHDKDLKVNSYLSKGLFERNYNGHGMEMAYNE 1530
 QY 1827 AMADALERFG-NRFVSLSDVLYGRNLFAMAVGVFELTYLLOYRFPTRPRVNAKL 1885
 DB 1531 YINEYVARIQGFQDKMSPEFWDIYTRGLVMAVGVGFLLTITMCQYNFLRQRMPEVST 1590
 QY 1886 SPLNDEDEVERERQRIIDGGGONDILEIKELTKTYRK--RRPAYDRICVGPPEECF 1942
 DB 1591 KPVND-DVDVASEQORVLRGADNDWAKTEMLTKYVSKRIGRLADRLCIGVRECEP 1649
 QY 1943 GLLVNAGKSTFKMLTGDTTVTRGDAFLNKSILSNIEHVHONMGVCPQFDATITELL 2002
 DB 1650 GXLGVNAGKSTFKMLTGDESTTGGEAFVNGHSLVELKIQVOQOSIGYCQOCALDELFT 1709
 QY 2003 GREHVEFPALLRGVPEKEVGVGEMAIKGLVYKYGKAYAGNISGCKRKLSTAMALIGG 2062
 DB 1710 AREHLQLTRRGIXMDEARVMALEKLELTKYADKPAAGTYSGGKRLSTALIALIG 1769
 QY 2063 PPVVFDEPTTGMPPKARRELMNLCALSVKREGRSVYLTSSMECEALCTRMIMVNGRP 2122
 DB 1770 PAFIFLDEPTTGMPPKARRELMNLDLITGSGSVYLTSSMECEALCTRLIMVNGRP 1829
 QY 2123 RCLGSVOHLKNRGCGTITIVRIAGSNPDLKPVQDFEGIALPGPSVLKEKHNNMLQYOLPS 2182

```
Dd 1830 RCLSTIOHLKRFEDGIMLIVR-TRKSQSKYDVRFRFNPNFEMAKLEHNHKVQYLKS 1888
      |||||:|||||::|||:~::~|||:|||||:|||||:
Oy 2183 SLSELIARFIISLSQSKRRLLIEDYSVSQTLLDOVFVNPAKSDSD 2227
      |||||:~::~|||:|||||:|||||:|||||:
Db 1889 EHSLAOYFSKMEQSVGLIEDIYSVSQTLIDNVFNPAKKQSDN 1933
      |||||:~::~|||:|||||:|||||:|||||:
```

```

RESULT 14
US-09-995-542-8
: Sequence 8, Application US/09995542
: Patent No. US20020127647A1
: GENERAL INFORMATION:
: APPLICANT: Shutter, John
: APPLICANT: Ullas, Iearn
: TITLE OF INVENTION: ATP-Binding Cassette Transporter-like Molecules and
: TITLE OF INVENTION: Uses Thereof
: FILE REFERENCE: 00-658-A
: CURRENT APPLICATION NUMBER: US/09/995,542
: CURRENT FILING DATE: 2001-11-28
: PRIOR APPLICATION NUMBER: 60/233,520
: PRIOR FILING DATE: 2000-11-28
: NUMBER OF SEQ. ID NOS: 24
: SOFTWARE: PatentIn Ver. 2.0
: SEQ. ID NO. 8
: LENGTH: 1550
: TYPE: prt
: ORGANISM: Homo sapiens
US-09-995-542-8

```

QY	829	LEFLELQWMTWYI EAVPEQOYI IPRPWY PCMCYMPGCEDEKSHSGSMQKRMSELCM	888
Db	600	LLDALALGLTWTLEAVCBOYI IPRPWYPRRRKSTWCGPRPKRPAACPPRLD -PKVLY	658
QY	889	EEEPYHLKLGAVSIQNLVKVYRDGMKVAVDGLALNFEQOITSFGLHNGAGKTTMSILTG	948
Db	659	EEAPRGISPCGVASVLEKPRPGSPQALRGSLIDFQOCHITAFGLHNGAGKTTTSLISG	718
QY	949	LEPPTSGATILCKDI RSEMSITRQMLGVCPOHNVLFQMLVMEHWPYARLKGISEKHV	1008
Db	719	LEPPSGASAPFLGDHVVSSMAALRPHLQVCPQNLVDFMLYDEHVVYGRYKLGISAAVY	778
QY	1009	KAEEMOALBYLJPRSSKRLSKTQSLQSGMQRLISVALAFVSGSKVYLIDEPTAGVPSYR	1068
Db	779	GPEDRRLIDQVL -VSKQSQYQTRHLSGQMQRKLSVAIAFVGSGQYVILDEPTAGVAPSR	837
QY	1069	RGIMELLKTRQCKTILTLSTHNDDEADYLDQRLAITSBKRLCCVSSSLFKMQJCTGYL	1128
Db	838	RGIMELLKTRQCKTILTLSTHNDDEADYLDQRLAIVAAVAGRLCCGSPFLFRHRLHSGYLL	897
QY	1129	TLVKKDVSSLSSCRNSSYVSLKKEDEVSSQSGSSDAGLSGSHESPDTLIDVSAISNLR	1188
Db	898	TLVKKARLPJLTN-----EKADTMESGVDTRQEKKNQSGHYGCPQALLAVQ	945
QY	1189	KHNSBARLVEDIGHETLYVLYPEAKKGAVALFPHEDIRLSDLGSSVGISETLLEIF	1248
Db	946	HWVGARLVELEPHLEVLVLYPYTGAMHGSFATLPFRLDRLAERLTGYSIGDSTLEIF	1005
QY	1249	LKVAEESVDJEDTDGJLPARRNRRAFQ -KQSCLPPTEDDAAPDNDSDIDPESREDL	1307
Db	1006	LKYVEECAADIDMDMGSCGQHLCTGJAGLDVYLRLLKMPQETALE -NCPAGSAPETQD	106
QY	1308	LSGMDKGSYQVKGWKLTOQOYVALMLKRLILRRSRKGFPAQIYLPANVYCALVFSLI	136
Db	1064	GSGDPAVQ -VYQGWALLRQDQALMLKRELLARSRRGIFAQIYLPALFVGLALVFSLI	1122
QY	1366	VPPFKYPSLELOPMWYNEQYTFVSNAPEDTQTELLNLTADPQFGTRCMGDNPIDT	1427
Db	1122	VPPFHYHIALSLPTMGVAQVSPFSEDAQPDPRARLLEALHQEAG-----LEBP	1177
QY	1428	PCQAGEEWTAPVRYQITMDLFQNGNTMOPSPBACCQSDKTKKMLPVCPRGAGLPP	148
Db	1172	PYQHSRHRFSAPRYPAEYAKVLASGNWTPESPSPACQCSOPGARLLPDCPPAAAGPRP	123
QY	1488	OKQOMTALIIDLDGRNLSIDVLYTYOYLAKSLKKIMWNERRYGSGFLGYSNFOALP	154
Db	1232	QAVTSGSEGVNLTGRNLSDFLVYTYRVLRYQCLTKIKWYNEVRYGSGFLG -GRDGLPS	129
QY	1548	SOEVDNAIKQMKHKLKLAQSSADRFNLISGRFMTGIDTRNNYKVPNNKGMAHAISSPLN	1607
Db	1291	QGELEGRSVEELMALISPLPGGALDRVLKNTLVANHSIDQADQSLIKIMFNKNGHSMVAEYN	1350
QY	1608	VYNNALILANLOKQENPSHVGITGFENPRLNLTQOJSEVALNMTASDVLYVSTICVJPMFS	166
Db	1351	RASNMYLKRHLRPGFARAHASTITLNLNRLNLTKQJSEALNMSVDVLYVSTICVJPMFS	1410
QY	1668	VPAFVEFVLIQERYSKAHNLOFISGVAPVYWLNSNFWDMQCNVBDATLVIIIFLICPOQ	1727
Db	1411	VPAFVLYVLIBRYBRAKHQJLMGSLPTLYWLGNFJLMDQCNLYVBDATLVIIIFLAFQOR	1476
QY	1728	SVYSSNTNPFVALLLLLYQWSTPFL 1752	
Db	1471	AVVAPRNPJALLLELLYGRQAVPY 1495	

```

RESULT 15
US-09-767-870-9
; Sequence 9, Application US/09767870
; Patent No. US20020037549A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: ABC Transport Polynucleotides, Polypeptides, and Antibodies

```

us-09-595-526c-2.rapb

```

1 FILE REFERENCE: P7010P1
2 CURRENT APPLICATION NUMBER:
3 PRIOR FILING DATE: 2001-01-24
4 PRIOR APPLICATION NUMBER: PCT/US00/19736
5 PRIOR FILING DATE: 2000-07-20
6 PRIOR APPLICATION NUMBER: 60/145,215
7 PRIOR FILING DATE: 1999-07-23
8 PRIOR APPLICATION NUMBER: 60/149,445
9 PRIOR FILING DATE: 1999-08-18
10 NUMBER OF SEQ ID NOS: 60/144,730
11 SOFTWARE: PatentIn Ver. 2.0
12 SEQ ID NO 9
13 LENGTH: 664
14 TYPE: PRT
15 ORGANISM: Homo sapiens
16 US-09-767-870-g

```

Every Match
Best Local
Match

1604 SFLN

1664 AMSEV

1724 FQOKSY

[illegible]

LSWDLVGRN

03 00-1111

1 VTRGDAFLNK
:11:1
3 Laspo-

SWAIRKGLV
TAGS: 1111

NSLAW

PAARS--OP

ESVSQTMLEFV

662

March 7, 1964

1

